

# Mapping the encounter state of a transient protein complex by PRE NMR spectroscopy

Alexander N. Volkov · Marcellus Ubbink ·  
Nico A. J. van Nuland

Received: 25 August 2010 / Accepted: 28 September 2010 / Published online: 4 November 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Many biomolecular interactions proceed via a short-lived encounter state, consisting of multiple, lowly-populated species invisible to most experimental techniques. Recent development of paramagnetic relaxation enhancement (PRE) nuclear magnetic resonance (NMR) spectroscopy has allowed to directly visualize such transient intermediates in a number of protein-protein and protein-DNA complexes. Here we present an analysis of the recently published PRE NMR data for a protein complex of yeast cytochrome *c* (Cc) and cytochrome *c* peroxidase (CcP). First, we describe a simple, general method to map out the spatial and temporal distributions of binding geometries constituting the Cc-CcP encounter state. We show that the spatiotemporal mapping provides a reliable estimate of the experimental coverage and, at higher coverage levels, allows to delineate the conformational space sampled by the minor species. To further refine the encounter state, we performed PRE-based ensemble simulations. The generated solutions reproduce well the experimental data and lie within the allowed regions of the

encounter maps, confirming the validity of the mapping approach. The refined encounter ensembles are distributed predominantly in a region encompassing the dominant form of the complex, providing experimental proof for the results of classical theoretical simulations.

**Keywords** Encounter state · Paramagnetic relaxation enhancement · Cytochrome *c* · Cytochrome *c* peroxidase · Transient complex

## Introduction

Many biomolecular interactions proceed via lowly populated, transient intermediates, which are increasingly recognized as important determinants of macromolecular recognition and association kinetics (Schreiber et al. 2009; Ubbink 2009). Due to a low population and inherent dynamics, this minor species is invisible to most structural and biophysical methods, which severely thwarts its experimental characterization. Recent advances in paramagnetic relaxation enhancement (PRE) nuclear magnetic resonance (NMR) spectroscopy have enabled direct visualization of transient intermediates in protein-protein (Tang et al. 2006; Volkov et al. 2006; Bashir et al. 2010; Xu et al. 2008) and protein-DNA complexes (Iwahara and Clore 2006) and biomolecular self-association (Tang et al. 2008a, b; Hartl et al. 2010). These and other studies (reviewed in refs. Schreiber et al. 2009; Ubbink 2009) have confirmed a long-held view that formation of a protein complex proceeds via a short-lived encounter state, which enables proteins to undergo reduced-dimensionality search of the optimal binding geometry, thereby accelerating molecular association as compared to 3D diffusion (Adam and Delbruck 1968).

**Electronic supplementary material** The online version of this article (doi:10.1007/s10858-010-9452-6) contains supplementary material, which is available to authorized users.

A. N. Volkov · N. A. J. van Nuland  
Department of Molecular and Cellular Interactions,  
VIB, Pleinlaan 2, 1050 Brussels, Belgium

A. N. Volkov (✉) · N. A. J. van Nuland  
Jean Jeener NMR Centre, Structural Biology Brussels,  
Vrije Universiteit Brussel, Pleinlaan 2,  
1050 Brussels, Belgium  
e-mail: ovolkov@vub.ac.be

M. Ubbink  
Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden  
University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

Recent experimental work on weak, transient protein interactions has revealed that the population of the encounter state—defined as percentage of time spent in this state relative to the total lifetime of the complex—varies in a wide range, spanning predominantly single-orientation systems (Tang et al. 2006; Volkov et al. 2006; Bashir et al. 2010) and highly dynamic, pure encounter complexes (Worrall et al. 2002; Xu et al. 2008). Moreover, it was shown that the amount of the encounter state in a protein–protein complex can be broadly modulated by interfacial point mutations (Volkov et al. 2010), suggesting an intriguing possibility of adjusting the population of the minor species.

PRE is caused by magnetic dipolar interactions between a protein nucleus and unpaired electrons of a paramagnetic probe (Clare 2008; Clare and Iwahara 2009), which can be introduced into the molecular frame by bioconjugation techniques. Due to the large magnetic moment of the unpaired electron and  $\langle r^{-6} \rangle$  distance dependence, protein nuclei located close to the paramagnetic center experience very large PREs, so that even lowly populated species can give rise to a measurable effect. This exquisite sensitivity makes PRE NMR spectroscopy a suitable tool for the study of transient intermediates in biomolecular interactions. For protein complexes in the fast exchange regime, the measured PRE is a population-weighted average of the contributions from all protein–protein orientations and, as such, contains the information on both the specific binding form and the encounter state (Clare 2008). In principle, PREs contain both temporal (population) and spatial (distances from the paramagnetic center) information on the minor species, which in favorable cases can be decomposed into separate contributions.

One of the systems studied by PRE NMR spectroscopy is a complex of cytochrome c (Cc) and cytochrome c peroxidase (CcP). Both proteins come from the intermembrane space of yeast mitochondria, where CcP catalyses the reduction of peroxides using the electrons donated by Cc—an important process mitigating the oxidative stress (Chance et al. 1967). In our earlier work we showed that interaction between Cc and CcP comprises a well-defined Cc–CcP form and a combination of non-specific protein–protein orientations (Volkov et al. 2006). The latter constitute an encounter state with the total population of 30% (Bashir et al. 2010). Here we present an analysis of the recently published, extended PRE dataset (Bashir et al. 2010). First, we describe a simple, general spatiotemporal mapping approach that provides a reliable estimate of the experimental coverage and, at higher coverage levels, allows to delineate the conformational space sampled by the minor species. Further, we use PRE-based ensemble simulations to refine the encounter state and show that

encounter ensembles are distributed predominantly in a region encompassing the dominant form of the complex, providing experimental proof for the results of classical theoretical simulations (Northrup et al. 1988). The combination of the methods used here is superior to a low-resolution, geometric analysis employed in our earlier work (Volkov et al. 2006) and offers a detailed visualization of the encounter state.

## Materials and methods

### Encounter state PREs

The transverse paramagnetic relaxation enhancement,  $\Gamma_2$ , is given by the Solomon–Bloembergen equation (Eq. 1; Solomon 1955; Solomon and Bloembergen 1956):

$$\Gamma_2 = \frac{1}{15} \left( \frac{\mu_0}{4\pi} \right)^2 \gamma_1^2 g^2 \mu_B^2 S(S+1) r^{-6} \left( 4\tau_c + \frac{3\tau_c}{1 + \omega_h^2 \tau_c^2} \right) \quad (1)$$

where  $r$  is the distance between the paramagnetic center and the observed proton,  $\mu_0$  is the permeability of vacuum,  $\gamma_1$  is the proton gyromagnetic ratio,  $g$  is the electron g-factor,  $\mu_B$  is the electron Bohr magneton,  $S$  is the electron spin number,  $\tau_c$  is the rotational correlation time, and  $\omega_h$  is the proton Larmor frequency. The rotational correlation time is defined as  $\tau_c = (\tau_r^{-1} + \tau_s^{-1})^{-1}$ , where  $\tau_r$  is the rotational correlation time of the protein complex (equal to 16 ns for Cc–CcP; Volkov et al. 2006) and  $\tau_s$  is the effective electron relaxation time. For a nitroxide SL used in this work,  $\tau_s \gg \tau_r$  so that  $\tau_c \approx \tau_r$  (Clare and Iwahara 2009).

For each Cc backbone amide ( $i$ ), the observed ( $\Gamma_{2,i}^{\text{obs}}$ ) is the sum of the population-weighted contributions of the specific form ( $\Gamma_{2,i}^{\text{specific}}$ ) and the encounter state ( $\Gamma_{2,i}^*$ ):

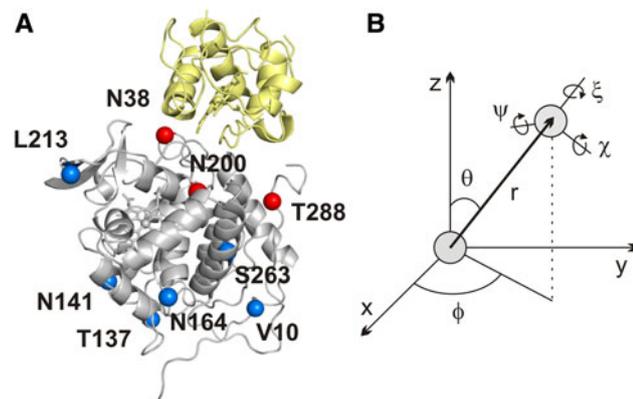
$$\Gamma_{2,i}^{\text{obs}} = p_{\text{tot}} \Gamma_{2,i}^* + (1 - p_{\text{tot}}) \Gamma_{2,i}^{\text{specific}} \quad (2)$$

where  $p_{\text{tot}}$  is the total population of the encounter state, defined as the percentage of time spent in this state relative to the total lifetime of the complex. The  $\Gamma_{2,i}^{\text{specific}}$  values were back-calculated from the crystal structure of the complex (PDB 2PCC; Pelletier and Kraut 1992) using prePot module (Iwahara et al. 2004) in Xplor-NIH (Schwieters et al. 2003, 2006). To account for the mobility of the attached SL, the calculated effects were averaged over an ensemble of 150 SL conformers generated by simulated annealing in torsion angle space (Iwahara et al. 2004). For the Cc residues exhibiting no PREs (i.e.  $I_{\text{para}}/I_{\text{dia}} > 0.8$ , where  $I_{\text{para}}$  and  $I_{\text{dia}}$  are peak intensities in the HSQC spectra of the spin-labeled complex and a diamagnetic control, respectively;

Bashir et al. 2010),  $\Gamma_{2,i}^{\text{obs}}$  were set to  $5 \text{ s}^{-1}$  and the errors derived from  $I_{\text{para}}/I_{\text{dia}}$  values as reported before (Bashir et al. 2010). Otherwise  $\Gamma_{2,i}^{\text{obs}}$  and their errors were taken from previous work (Bashir et al. 2010). For each of the 10 SL conjugation sites, the  $\Gamma_{2,i}^*$  values were obtained from Eq. 2 and used in further analysis.

### Generating the conformational space grid

All molecular simulations were performed in Xplor-NIH (Schwieters et al. 2003, 2006). The coordinates of Cc–CcP complex were taken from the X-ray structure (PDB 2PCC; Pelletier and Kraut 1992) and oriented such that centers of mass (CMs) of CcP and Cc appeared at the origin of the coordinate system and on the positive  $z$  axis, respectively. The position of CcP was fixed, while Cc molecule was systematically rotated around  $x$  and  $z$  axes, corresponding to  $\theta$  and  $\varphi$  rotations around CcP in the spherical coordinate space (Fig. 1b). The rotation increments  $\delta\theta$  and  $\delta\varphi$  determine the desired spatial resolution, which in our case was set to 1 Å separation between neighboring Cc CMs. To emulate the rotational freedom, Cc was rotated around orthogonal  $\chi$ ,  $\psi$ ,  $\xi$  axes originating at its CM (Fig. 1b). By systematically varying the rotational coordinates  $\chi$ ,  $\psi$ ,  $\xi$  ( $0 \leq \chi \leq 2\pi$ ,  $0 \leq \psi \leq 2\pi$ ,  $0 \leq \xi < \pi$ ) in the increments of  $\delta\chi = \delta\psi = \delta\xi = \pi/3$ , a set of 108 non-redundant Cc rotamers was produced at each  $(\theta, \varphi)$  position. For every



**Fig. 1** Specific form of Cc–CcP complex. **a** Crystallographic Cc–CcP binding orientation (Pelletier and Kraut 1992). Cc and CcP are in grey and yellow, with heme groups in sticks. Cz atoms of CcP residues used for spin-labeling are shown as spheres, colored according to whether the attached SL exhibits intermolecular PREs (red) or not (blue). One SL position (K97, blue) is not seen in this view. **b** Definition of the spherical coordinates used in this work. Proteins' centers of mass are shown as grey spheres, with CcP at the origin of the coordinate system. Cc possesses three translational ( $\theta$ ,  $\varphi$ ,  $r$ ) and three rotational ( $\chi$ ,  $\psi$ ,  $\xi$ ) degrees of freedom. For the definition of the rotational axes see “Materials and methods”. All protein representations in this work are visualized with PyMOL (DeLano 2002)

( $\theta$ ,  $\varphi$ ,  $\chi$ ,  $\psi$ ,  $\xi$ ) combination, the intermolecular van der Waals (vdW) energy term was calculated, with vdW potential set to zero for protein sidechain atoms extending beyond  $C\beta$ . Cc was then translated along the vector joining Cc and CcP CMs in steps of 1 Å until the vdW energies reached the values between zero and a chosen cut-off, thus either relieving intermolecular steric clashes or bringing together separated molecules in a rigid-body mimic of a protein complex. The distance between protein CMs at each  $(\theta, \varphi, \chi, \psi, \xi)$  defines the other translational coordinate,  $r$ . In this way, we explored the entire conformational space available to the interacting proteins ( $0 \leq \theta \leq \pi$ ,  $0 \leq \varphi < 2\pi$ ), sampling 12,205  $(\theta, \varphi)$  positions and producing a total of 1,318,140 Cc–CcP orientations at varying  $(\theta, \varphi, r, \chi, \psi, \xi)$ .

### Mapping the encounter state

For each of 1,318,140  $(\theta, \varphi, r, \chi, \psi, \xi)$  orientations, the expected PREs ( $\Gamma_{2,i}$ ) were back-calculated as described above, and the maximal population ( $p_{\text{max}}$ ) at which no violations of the experimental  $\Gamma_{2,i}^*$  restraints occurred was obtained (Eq. 3):

$$p_{\text{max}} \begin{cases} \min_i \left( \Gamma_{2,i}^* / \Gamma_{2,i} \right), & \min_i \left( \Gamma_{2,i}^* / \Gamma_{2,i} \right) < p_{\text{tot}} \\ p_{\text{tot}}, & \min_i \left( \Gamma_{2,i}^* / \Gamma_{2,i} \right) \geq p_{\text{tot}} \end{cases} \quad (3)$$

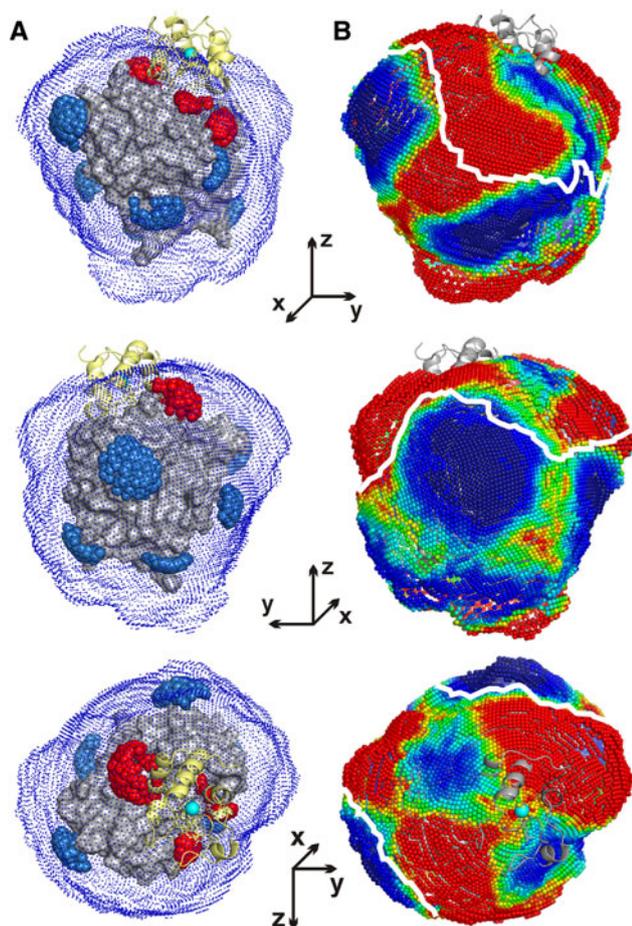
To visualize the results, the largest  $p_{\text{max}}$  of 108  $\chi, \psi, \xi$  Cc rotamers at each  $(\theta, \varphi)$  position [ $p_{\text{max}}(\theta, \varphi)$ , Eq. 4] was noted, and  $p_{\text{max}}(\theta, \varphi)$  values were color-coded onto the interaction grid isosurface  $\theta, \varphi, r_{(\chi, \psi, \xi=0)}$ , thus producing the spatiotemporal map shown in Fig. 2b.

$$p_{\text{max}}(\theta, \varphi) = \max_{\chi, \psi, \xi} [p_{\text{max}}(\chi, \psi, \xi)_{\theta, \varphi = \text{const}}] \quad (4)$$

To delineate the area containing protein–protein orientations contributing to  $\Gamma_2^*$  (restricted by the white curve in Fig. 2b), we defined a set of encounter PRE restraints for Cc residues that exhibit violations of  $\Gamma_{2,i}^{\text{obs}}$  in the specific Cc–CcP complex (i.e. highlighted areas in Fig. 4c) with  $(\Gamma_{2,i}^{\text{obs}} - \delta\Gamma_{2,i}^{\text{obs}}) - \Gamma_{2,i}^{\text{specific}} > 5 \text{ s}^{-1}$ , where  $\delta\Gamma_{2,i}^{\text{obs}}$  is the error of  $\Gamma_{2,i}^{\text{obs}}$ , and selected all Cc molecules that contribute at least 5 Hz to these encounter restraints at a given  $p$ . The scripts for the encounter state mapping are provided in Supplementary Material.

### Ensemble refinement against intermolecular PREs

Using the  $\Gamma_2^*$  dataset obtained from all 10 SL conjugation sites (see above), the rigid-body simulated annealing refinement of the Cc–CcP encounter state was carried out in Xplor-NIH (Schwieters et al. 2003, 2006) following the



**Fig. 2** Spatiotemporal analysis of Cc–CcP encounter state. **a** Interaction grid isosurface  $\theta, \varphi, r_{(x,\psi,\xi=0)}$  consisting of 12,205 Cc CMs (blue dots). **b** Isosurface in **(a)** coloured according to  $p_{\max}(\theta, \varphi)$  (Eq. 4), ranging from 0 (blue) to 0.3 (red). The white curve limits an area around the dominant form of the complex that contains Cc orientations contributing  $\geq 5$  Hz to  $\Gamma_{2,i}^*$ . CcP (grey surface) and Cc (cartoon) in the top panels are in the same orientation as in Fig. 1a. The middle and bottom views are obtained by 180° rotation of the top-panel representations around, respectively, z and y axes. For each SL the oxygen atoms of 150 conformers used for ensemble averaging are space filled and colored red and blue to indicate, respectively, the presence and absence of the measured paramagnetic effects. The cyan sphere shows Cc CM in the dominant complex

published procedure (Tang et al. 2006). Briefly, the position of CcP was fixed, and multiple copies of Cc molecules, representing ensembles with  $N = 1–20$ , were docked to minimize the energy function consisting of the PRE target term, vdW repulsion term to prevent atomic overlap between Cc and CcP, and a weak radius-of-gyration restraint used to encourage intermolecular Cc–CcP contacts (Tang et al. 2006). Note that this procedure allows for the atomic overlap among Cc molecules constituting an ensemble. As a rule, 100 independent refinement runs were performed.

To assess the agreement between the observed PREs and the PREs back-calculated from Cc ensembles generated in each run, we calculated a  $Q$  factor (Eq. 5):

$$Q = \sqrt{\frac{\sum_j \sum_i (\Gamma_{2,ij}^{\text{obs}} - \Gamma_{2,ij}^{\text{calc}})^2}{\sum_j \sum_i (\Gamma_{2,ij}^{\text{obs}})^2}} \quad (5)$$

where  $j = 1–3$  runs over three SL positions showing paramagnetic effects (N38C, N200C and T288C; Fig. 1a) and  $\Gamma_{2,i}^{\text{calc}}$  is given by Eq. 6:

$$\Gamma_{2,ij}^{\text{calc}} = \frac{p_{\text{tot}}}{N} \sum_{k=1}^N \Gamma_{2,ijk}^* + (1 - p_{\text{tot}}) \Gamma_{2,ij}^{\text{specific}} \quad (6)$$

where  $p_{\text{tot}}$  is the total population of the encounter state,  $N$  is the size of the encounter ensemble,  $\Gamma_{2,ijk}^*$  is the PRE from SL ( $j$ ) back-calculated for the residue ( $i$ ) of the Cc ensemble member ( $k$ ), and  $\Gamma_{2,ij}^{\text{specific}}$  is the PRE back-calculated from SL ( $j$ ) for the residue ( $i$ ) of Cc in the dominant form of the complex. The reported  $Q_e$  is the average  $Q$  factor obtained from the ensembles generated in repeated refinement runs, while  $Q_{ec}$  is the ‘ensemble of ensembles average’ (Tang et al. 2006) calculated by using the average  $\Gamma_{2,ij}^{\text{calc}}$  computed from all  $n$  ensembles (Eq. 7):

$$\Gamma_{2,ij}^{\text{calc}} = \frac{p_{\text{tot}}}{nN} \sum_{m=1}^n \sum_{k=1}^N \Gamma_{2,ijkm}^* + (1 - p_{\text{tot}}) \Gamma_{2,ij}^{\text{specific}} \quad (7)$$

## Results

### Mapping the encounter state

To map out the conformational space occupied by the Cc–CcP encounter state, we have analyzed the intermolecular paramagnetic effects exerted on Cc nuclei by an unpaired electron of a nitroxide spin-label (SL) placed at ten different positions, one at a time, on the surface of CcP (Fig. 1a). As reported before (Volkov et al. 2006; Bashir et al. 2010), three SLs located close to the crystallographic binding site (N38C, N200C and T288C, shown as red spheres in Fig. 1a) give rise to PREs, while SLs attached to any of the other seven positions (blue spheres in Fig. 1a) show no effects. Most of the observed PREs arise from the dominant form of Cc–CcP complex; however, several Cc regions experience additional paramagnetic effects (highlighted in Fig. 4c), originating from protein–protein orientations constituting the encounter state (Volkov et al. 2006). By subtracting the effects of the dominant orientation from the observed PREs (Eq. 2 in “Materials and methods”), we obtained a set of the encounter state’s PRE contributions ( $\Gamma_{2,i}^*$ ). This dataset, together with the

information from the SLs exhibiting no measurable effects, was used in the subsequent analysis.

First, using a rigid-body sampling procedure (see “Materials and methods”), we generated a grid of Cc–CcP orientations, corresponding to the entire conformational space available to the interacting proteins (Fig. 2a). Note that each dot in Fig. 2a represents the centre of mass (CM) of Cc orientations with the same  $(\theta, \varphi)$  coordinates, produced by non-redundant rotations around  $\chi, \psi, \xi$  axes (see Fig. 1b for axes definition and “Materials and methods” for details). Second, for each of the generated orientations, we back-calculated the expected PREs ( $\Gamma_{2,i}$ ) and obtained the maximal population ( $p_{\max}$ ) at which no violations of the experimental  $\Gamma_{2,i}^*$  restraints occurred (Eq. 3). Finally, for each grid point in Fig. 2a, the largest  $p_{\max}$  of  $\chi, \psi, \xi$  Cc rotamers [ $p_{\max}(\theta, \varphi)$ , Eq. 4] was noted, and all  $p_{\max}(\theta, \varphi)$  values were color-coded onto the interaction isosurface, providing spatial (location) and temporal (population) map of the encounter state distribution (Fig. 2b and Supplementary Movie S1). The spatiotemporal map in Fig. 2b delineates the extent of the conformational space accessible to Cc–CcP orientations with populations  $p$ . In other words, the map shows regions of space where all solutions for a given  $p$  are to be found. This all-inclusiveness of  $p$  solutions is a salient feature of the spatiotemporal encounter maps and an important achievement afforded by the present approach.

However, there are several drawbacks associated with the current analysis. First, such ‘zero-resolution’ approach provides no molecular-level details on protein–protein geometries constituting the encounter state. Second, the introduced spatiotemporal maps outline the areas that *can be*, but not necessarily *are*, populated in the encounter state. Thus, to pin down the actual region occupied by an encounter ensemble, an adequate experimental coverage of the *entire* conformational space is essential. To illustrate this point, imagine that no experimental PRE data on Cc–CcP complex has yet been collected. Following our reasoning, the entire interaction surface in Fig. 2b can be painted red, with  $p_{\max} = p_{\text{tot}}$  for all grid points. In other words, with no a priori assumptions, encounter ensemble members can be located anywhere in the conformational space, and their populations range from zero to the total encounter population,  $p_{\text{tot}}$ . To continue our thought experiment, imagine that the first PRE dataset has been collected and, for simplicity’s sake, the introduced SL exhibited no paramagnetic effect. This would allow us to color an area next to the SL in blue, indicating that only protein–protein orientations with very low populations, if any, can be found there, thereby restricting the effective conformational space available to the encounter. Addition of more experimental data from SLs placed at other surface

positions would restrict the red area even further, bringing us closer to the actual region encompassed by the encounter state. Note that at this stage the SLs exhibiting no effects are as valuable as those showing PREs, because they allow for large portions of no-go space to be carved out. Ultimately, with an adequate experimental coverage, we end up with a warm-color area found only around the SLs showing  $\Gamma_2^*$  effects, which indicates the true location of the encounter space.

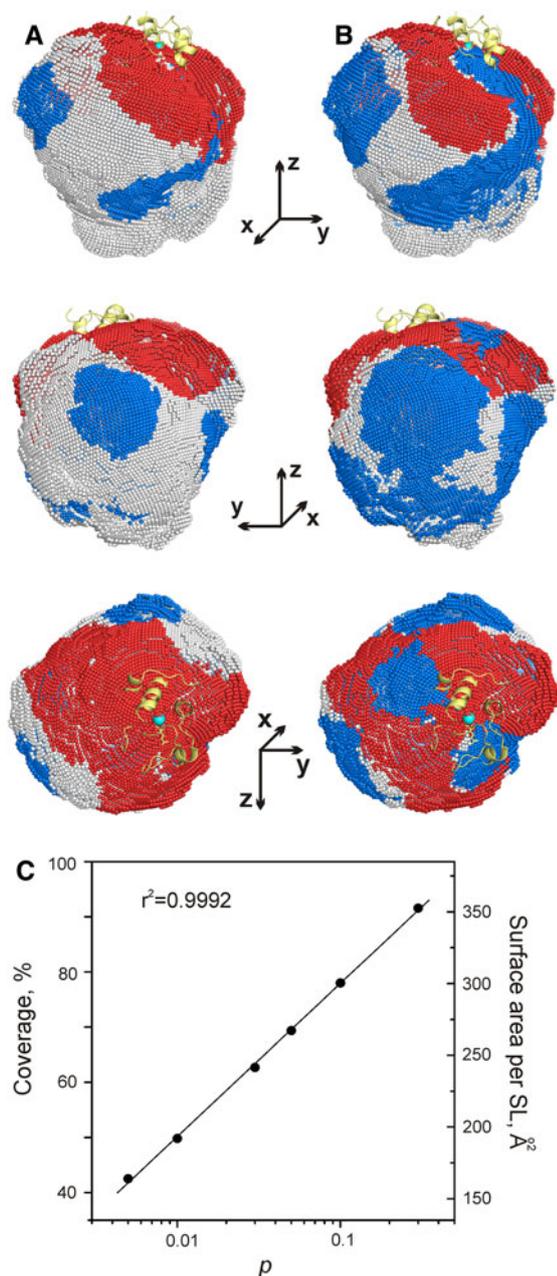
Going back to Fig. 2b, we notice that a large part of the isosurface is composed of warm-color grid points. Most of these correspond to Cc molecules that contribute to the experimental  $\Gamma_2^*$  restraints (the region above the white curve in Fig. 2b), thus defining the extent of the encounter space at the current level of experimental coverage. However, many warm-color points lie outside this area, which indicates incomplete PRE sampling. To assess the experimental coverage of the Cc–CcP encounter state, we mapped out the regions containing Cc molecules contributing to (red) or violating (blue) the experimental  $\Gamma_2^*$  restraints at different  $p$  values (Fig. 3a, b and Supplementary Fig. S1). In these views, the conformational space not covered by the effects from the introduced SLs is shown in white. As expected, the white areas increase with decreasing  $p$  values, implying that progressively more experimental input is required to track down more lowly populated species.

Integration over red and blue areas in Fig. 3a, b provides a simple means of quantifying the extent of the experimental coverage. In our case, there is a good, log-scale correlation between  $p$  and the calculated coverage (Fig. 3c). Thus, about one half of the conformational space is probed by PREs at  $p = 0.01$ , increasing to nearly 80% at  $p = 0.1$  (Table 1, cf. Fig. 3a, b). We estimate that 13–20 SLs per CcP—corresponding to one SL attached per each 190–300 Å<sup>2</sup> of the total surface area—are required to provide an adequate coverage at  $p = 0.1$ –0.01 (Table 1).

#### Ensemble refinement of the encounter state

Direct use of  $\Gamma_2^*$  restraints in an ensemble-based, rigid-body simulated annealing structure calculation protocol—pioneered by Clore and co-workers (Iwahara et al. 2004; Schwieters et al. 2006; Tang et al. 2006)—provides an alternative, potentially more informative, means of refining the encounter state. In this approach, multiple copies of Cc are docked simultaneously to CcP by minimizing the difference between the combination of PREs from all Cc molecules and the experimental  $\Gamma_2^*$  values (see “Materials and methods” for details).

We performed multiple structure calculations with ensemble sizes ( $N$ ) varying from 1 to 20 and the encounter



**Fig. 3** Experimental coverage of the conformational space of Cc–CcP encounter state. Cc CMs for the orientations that, respectively, violate experimental PREs (*blue*) or contribute  $\geq 5$  Hz to  $\Gamma_2^*$  (*red*) at **a**  $p = 0.01$  and **b**  $p = 0.1$ . Combination of the *blue* and *red* areas defines the total conformational space covered by the effects from the introduced SLs. The same isosurface as in Fig. 2a, b is shown. See the legend to Fig. 2 for more details. **c** Plot of experimental coverage, or surface area of CcP covered by PREs from a single SL, versus the population of the minor species

state population  $p_{\text{tot}} = 0.3$ , determined in our earlier work (Bashir et al. 2010, also see below). To assess the quality of solutions, we calculated a  $Q$  factor (Iwahara et al. 2003, 2004), which is a measure of agreement with the

**Table 1** Experimental coverage of the conformational space

$p$	Coverage (%)	Number of SLs <sup>a</sup>	Surface area (Å <sup>2</sup> ) per SL <sup>a,b</sup>
0.005	42.5	24	164 (138)
0.01	49.8	20	192 (162)
0.03	62.7	16	242 (204)
0.05	69.4	14	268 (226)
0.1	78.0	13	301 (254)
0.3	91.6	11	353 (298)

<sup>a</sup> Required for complete coverage

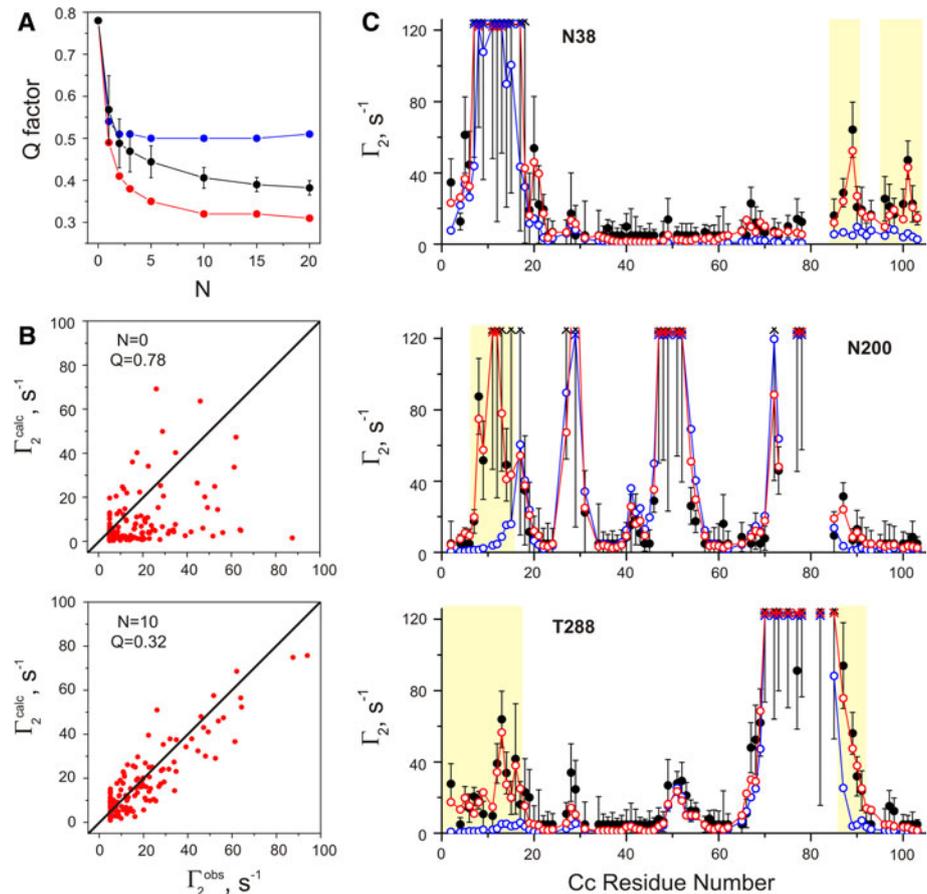
<sup>b</sup> Surface area excluding the binding site is given in parentheses

experimental data (the smaller the  $Q$  factor, the better the agreement; Eq. 5). In Fig. 4a,  $Q_e$  (an average  $Q$  factor of the individual ensembles) and  $Q_{ee}$  (a  $Q$  factor calculated by averaging PREs of Cc molecules in all ensembles; Tang et al. 2006) are plotted as a function of  $N$ . The  $Q$  factors diminish with the increasing ensemble size, leveling off at  $N = 10$ –20. As can be seen from Fig. 4a,  $Q_{ee}$  is systematically smaller than  $Q_e$ , which is due to the stochastic rather than unique combination of protein–protein orientations within each ensemble, such that averaging over all ensembles leads to a better agreement with the data (Tang et al. 2006). By randomly omitting 10% of  $\Gamma_2^*$  restraints and verifying how well these ‘free’ PREs are predicted by the remaining, ‘working’ data set (i.e. 90% included in the refinement), we performed a complete cross-validation (Brünger et al. 1993), with  $Q_{\text{free}}$  as a measure of the fit. The calculated  $Q_{\text{free}}$  values (Fig. 4a) indicate that  $N = 10$ –20 is the optimal size of the Cc ensemble required to satisfy the experimental restraints and that the improvement in the  $Q$  factors is not due to over-fitting (Tang et al. 2006).

As can be seen from  $\Gamma_2^{\text{obs}}$  vs.  $\Gamma_2^{\text{calc}}$  plots (Fig. 4b) and PRE profiles (Fig. 4c), a combination of PREs from the refined encounter ensemble and the dominant, crystallographic Cc–CcP orientation provides a good agreement with the experimental data. Clearly, most of the encounter PRE restraints are now satisfied (highlighted regions in Fig. 4c). To visualize the distribution of Cc molecules in the encounter state, we use a reweighted atomic probability density map (Schwieters and Clore 2002), calculated from 100 independently generated ensembles with  $N = 10$  (Fig. 5a). Most of the minor species are found in an area surrounding the dominant form of the complex, and a small, low-density patch of solutions is located at the back of CcP (see below).

Note that the atomic probability density maps in Fig. 5a are derived from all Cc atoms, while spatiotemporal maps in Figs. 2, 3 show only the CMs. Thus, to

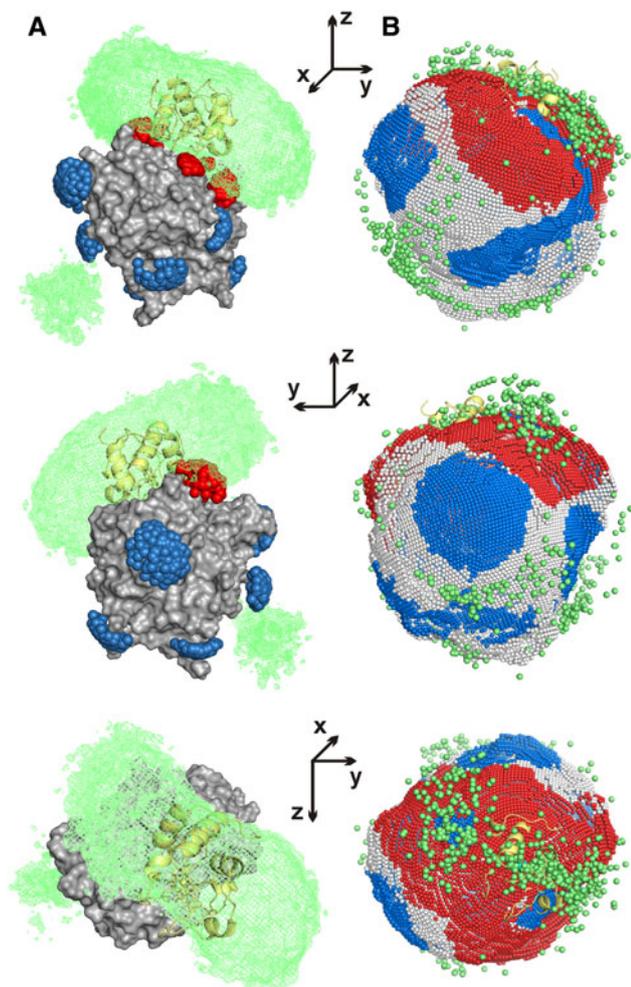
**Fig. 4** PRE-based ensemble simulations of the Cc–CcP encounter state. **a** Intermolecular  $Q$  factors:  $Q_e$  (black),  $Q_{ee}$  (red), and  $Q_{free}$  (blue). See text for the definitions. **b** Correlation between the observed ( $\Gamma_2^{obs}$ ) and calculated ( $\Gamma_2^{calc}$ ) PREs for the dominant form of the complex alone ( $N = 0$ , top) or in combination with the simulated encounter ensemble ( $N = 10$ , bottom). **c** Observed and calculated PREs for Cc–CcP–SL complexes, with SLs attached at position N38C (top), N200C (middle) and T288C (bottom). Experimental  $\Gamma_2^{obs}$  (black; Volkov et al. 2006; Bashir et al. 2010),  $\Gamma_2^{calc}$  for the specific orientation (blue), and  $\Gamma_2^{calc}$  for the combination of the specific form and an encounter ensemble ( $N = 10$ , red). Crosses indicate the value of  $\Gamma_2 \geq 125 \text{ s}^{-1}$  for the calculated PREs or identify the residues whose resonances disappear in the paramagnetic spectrum. The errors are standard deviations



enable a direct comparison of the two representations, we plotted the CMs of Cc molecules from 100 generated ensembles ( $N = 10$ ) together with the interaction grid isosurface contoured at  $p = p_{tot}/N = 0.03$  (Fig. 5b and Supplementary Movie S2). With a few exceptions, Cc CMs are located in the allowed regions of the encounter maps, indicating a good agreement between the two methods. It should be noted that, unlike in the spatio-temporal mapping approach, small violations are tolerated in the simulated annealing ensemble refinement: a slight violation of one restraint, accompanied by concomitant satisfaction of several others, can provide a better agreement with the experimental data than a good solution for the same restraint coming at a price of multiple bad solutions for others.

Despite making no individual contributions to the observed PREs, Cc molecules found in the white regions of the encounter maps nevertheless influence the ensemble-averaged  $\Gamma_2^*$  values obtained in the refinement procedure. The presence of such non-contributing solutions (e.g. in a low-density region at the back of CcP, Fig. 5a) could signify: (1) excessive ensemble size, (2) incorrect population of the encounter state used in the

calculations, or (3) insufficient experimental coverage of the conformational space. In our case, the first of these possible causes can be dismissed as decreasing the ensemble size from  $N = 10$  to  $N = 5$  to  $N = 3$  does not completely eliminate the non-contributing solutions and steadily increases the  $Q$  factor (Fig. 4a). Moreover, the complete cross-validation of  $\Gamma_2^*$  dataset ruled out a possible over-fitting at higher  $N$  values (see above). To test the second possibility, we repeated calculations at different  $p_{tot}$  for ensembles with  $N = 5$  and  $N = 10$  (Supplementary Fig. S2). In both cases, the  $Q$  factors fall sharply from  $p = 0$  to  $p = 0.3$  and then level off at  $p = 0.3$ – $0.5$ , confirming that the value of  $p_{tot} = 0.3$ , determined in a recent study (Bashir et al. 2010) and used throughout this work, is correct. Finally, to explore the third option, we performed control runs in which the number of Cc molecules in the ensemble was varied from  $N = 5$  to  $N = 9$  but their individual populations kept constant at  $p_i = 0.03$ , so that  $\sum_i p_i < p_{tot}$ . In this way, we assessed whether a subset of binding geometries with the combined population of  $0.15 \leq \sum_i p_i \leq 0.27$  can account for  $\Gamma_2^*$  effects of the entire encounter state ( $p_{tot} = 0.3$ ). In the control runs, decrease in  $\sum_i p_i$  is accompanied by only



**Fig. 5** Spatial distribution of the encounter ensembles. **a** Reweighted atomic probability density maps for the overall distribution of Cc molecules obtained from 100 PRE-based ensemble calculations ( $N = 10$ , plotted at a threshold of 20% maximum). In the *bottom view*, SL atoms are removed for clarity. **b** Overlay of the Cc–CcP interaction isosurface coloured according to the experimental PREs at  $p = 0.03$  (see the legend to Fig. 3 for details) and CMs of Cc molecules from 100 PRE-based ensemble simulations ( $N = 10$ , *green spheres*)

a small increase in  $Q$  factors (Supplementary Fig. S3), and the overall distribution of encounter ensembles remains essentially the same as that shown in Fig. 5a, except that the low-density patch at the back of CcP is steadily reduced with decreasing  $\Sigma_i p_i$  (e.g. compare the views for  $\Sigma_i p_i = 0.21$  in Supplementary Fig. S4 and  $\Sigma_i p_i = p_{\text{tot}} = 0.3$  in Fig. 5a). These results indicate that the experimental  $\Gamma_2^*$  values can be accounted for by a limited subset of protein–protein orientations, suggesting that Cc ensemble members found in the white regions of the encounter maps might represent a minor sub-population of the encounter state, not reported upon by the SLs introduced so far.

## Discussion

### Experimental description of the encounter state

The spatiotemporal mapping approach presented here is superior to a simple, geometric analysis of the encounter state employed in our earlier work (Bashir et al. 2010; Volkov et al. 2006) in that it uses protein structures and realistic van der Waals potentials, rather than spheres and uniform cut-off values, to sample the conformational space; relies on explicit  $\Gamma_2^*$  data, instead of uniform estimates, for calculation of allowed  $p$  values; and utilizes extensive ensemble-averaging of the PRE effects over multiple SL conformers, thus accounting for the mobility of the attached paramagnetic probes. When applied to an extended experimental dataset spanning 10 SL positions, these methodological advances result in a more informative and detailed encounter map compared to our earlier, roughly shaped “clouds” drawn from the effects of 5 SLs (Volkov et al. 2006).

The main advantage offered by the encounter maps is that they include all possible spatial solutions for  $Q \rightarrow 0$  at  $\Sigma_i p_i = p_{\text{tot}}$ . However, this comes at a price of providing no molecular-level details on the protein–protein orientations constituting the encounter state. To overcome the ‘zero-resolution’ limitation of the mapping, the encounter space was further refined by restrained ensemble simulations, affording a more detailed description of the minor species. The generated solutions reproduce well the experimental data (Fig. 3c); however, the  $Q$  factor ( $Q_{\text{cc}} = 0.32$  for  $N = 10$ –20) is slightly higher than those obtained in PRE NMR studies of other biomolecular interactions (Iwahara and Clore 2006; Tang et al. 2006). This can be attributed to large errors on the experimental  $\Gamma_2^{\text{obs}}$  values (Volkov et al. 2006; Bashir et al. 2010), obtained from intensity analysis of HSQC spectra (Battiste and Wagner 2000). In principle, longer spectral acquisition or the use of a two-point  $\Gamma_2$  measurement scheme (Iwahara et al. 2007; Clore and Iwahara 2009) could increase both accuracy and precision of the data. In practice, however, the instability of Cc–CcP complex—caused by autoreduction of Cc (Young and Caughey 1987) occurring on time scale of several hours (A.N.V., M.U. unpublished observations)—severely restricts the effective experimental time, precluding the use of two-point  $\Gamma_2$  measurements. Still, despite the practical limitations inherent in our system, the PRE NMR analysis provides a meaningful picture of the Cc–CcP encounter state.

There is a certain overlap between the concepts of the two approaches used here to analyze the encounter state. For instance, the number of Cc molecules included in the ensemble simulations could be thought of as defining the

size of a brush to paint the encounter space map, imbuing it with spatial resolution. Concerning the temporal resolution, though all ensemble members are uniformly populated at  $p = p_{\text{tot}}/N$ , allowing for the overlap of Cc molecules during the simulations effectively reproduces non-uniform populations captured in the encounter maps. The major conceptual difference between these methods is that the encounter mapping is inherently negative (or exclusive, i.e. relies on carving out the regions of space that *cannot* be populated at a given  $p$ ), while ensemble simulations are essentially positive (or inclusive, i.e. finding the solutions that *satisfy* given restraints). As a result, the former benefits from SLs exhibiting no PRE effects and is sensitive to the extent of the experimental coverage, while the latter relies on the observed PREs and is more tolerant to incomplete experimental sampling.

Narrower distribution of Cc CMs in the  $N = 10$  ensembles, compared to the red area of the encounter map (Fig. 5b), indicates that only a limited subset of allowed solutions has been found in the ensemble simulations. This can be due to an incomplete experimental coverage of the encounter maps or an insufficient sampling during the refinement procedure. The former can be improved by the introduction of more SLs to further restrict the encounter space, while the latter may be remedied by a more aggressive search. Alternatively, to tease out encounter ensembles directly from the spatiotemporal maps, one could sample multiple combinations of allowed orientations in search for the ones reproducing the experimental  $\Gamma_2^*$  data, using a suitable algorithm (e.g. a metaheuristic search; A.N.V. work in progress).

We would like to stress that the spatiotemporal map presents  $p_{\text{max}}$  values for individual Cc–CcP orientations, some of which could populate the encounter state (enclosed by the white curve in Fig. 2b). Reconstitution of the analogous map for the entire encounter state is a non-trivial, multivariate problem. For instance, it is conceivable that a combination of protein–protein orientations, each of which is allowed individually at a given  $p$ , will summarily yield a prohibitively high  $\Gamma_2^*$  value, violating the experimental PRE. Thus, a good sampling of multiple combinations of allowed orientations would be required to glean the total encounter state map. One way to approach this problem is offered by the metaheuristic search mentioned above, which is currently under investigation in our laboratory.

In its use of  $p_{\text{max}}$ —the maximal population of a particular orientation compatible with the experiment—the present approach is akin to the method of maximum allowed probabilities, developed to characterize flexible, partially independent protein domains from residual dipolar couplings and pseudocontact shifts (Gardner et al. 2005;

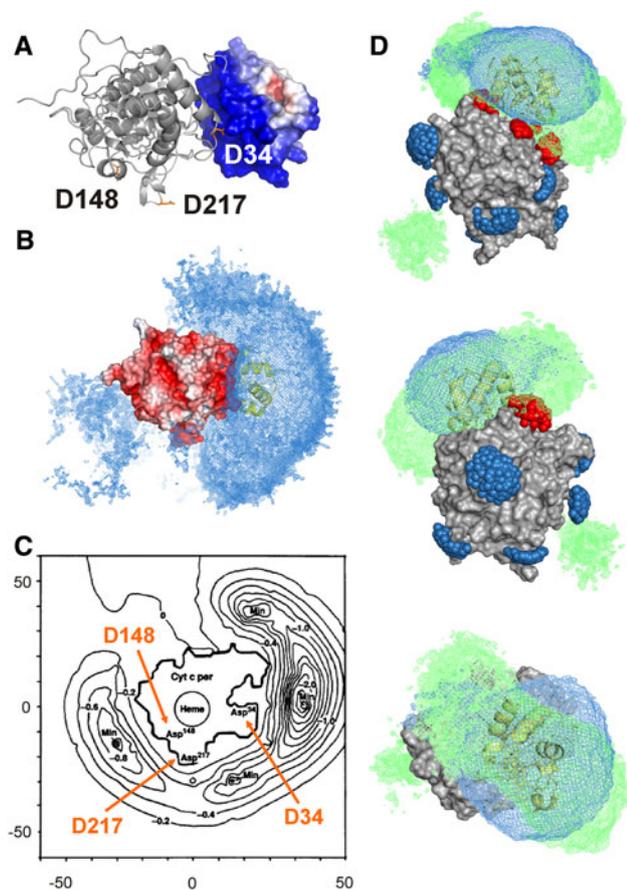
Longinetti et al. 2006; Bertini et al. 2007) and recently extended to small-angle X-ray scattering data (Bertini et al. 2010). Here we show that a similar idea can be successfully applied to characterization of protein–protein interactions by PRE NMR spectroscopy.

#### Comparison with theoretical simulations

The interaction between oppositely charged Cc and CcP was studied before by theoretical simulations employing Poisson–Boltzmann electrostatic potentials (Fig. 6a, c; Northrup et al. 1988; Gabdoulline and Wade 2001; Bashir et al. 2010). As shown in our recent work (Bashir et al. 2010), the ensemble of protein–protein geometries generated by electrostatics-based Monte Carlo (MC) protocol provides a good description of the Cc–CcP encounter state. In Fig. 6b, a typical MC ensemble is visualized using a reweighted atomic probability density map, revealing a good agreement with the results of a classical Brownian dynamics study (Fig. 6c; Northrup et al. 1988). In particular, four energy minima shown in the latter are also present in the MC simulations. In our case, the energy minimum around D148 is shallower, possibly due to the difference in the electrostatic potentials of Cc molecules used in the simulations (horse heart Cc in those of Northrup et al. 1988 and yeast *iso-1* Cc in our case).

The density maps generated from the theoretical (MC) and experimental (PRE-based) encounter ensembles encompass an area around the dominant form of Cc–CcP complex and broadly overlap (Fig. 6d). Despite a similar location, the MC and PRE ensembles exhibit different  $Q$ -factors [ $Q_{\text{cc}} = 0.54$  and  $0.32$  ( $N = 10$ ), respectively], indicating that the latter reproduce the experimental  $\Gamma_2^{\text{obs}}$  better. This is further evidenced by comparison of the corresponding PRE profiles (Fig. 4c here and Fig. 4a–c in Bashir et al. 2010). In this light, a somewhat broader distribution of the PRE ensembles could suggest that, in addition to electrostatics, other intermolecular forces may contribute to protein–protein interactions in the encounter state.

It should be noted that the MC solutions are the result of theoretical simulations, while the encounter ensembles described in this work are the product of the direct refinement against the measured PREs, which explains better agreement of the latter with the experimental data. Still, as shown in our earlier work (Bashir et al. 2010) and confirmed here by PRE-based ensemble refinement at different  $p$  values (see above), MC simulations provide a good representation of the encounter state and offer a robust estimate of its population. However, to obtain a more detailed description of the encounter state's conformational space, a further refinement of MC solutions appears to be necessary.



**Fig. 6** Comparison of the theoretical and experimental simulations of the Cc–CcP encounter state. **a** Crystallographic Cc–CcP orientation. To facilitate the comparison with the published data, three aspartates of CcP are labeled and shown as orange sticks. **b** Overall distribution of 1,701 Cc molecules in the simulated Cc–CcP encounter complex (Bashir et al. 2010), displayed as a reweighted atomic probability density map (Schwieters and Clore 2002; plotted at a threshold of 0.5% maximum, blue). The surfaces of **a** Cc and **b** CcP are colored by the electrostatic potential calculated at  $\pm 5 k_B T$  (red—negative, blue—positive) with APBS (Baker et al. 2001). **c** The Boltzmann-averaged total electrostatic potential energy of interaction between CcP and horse Cc in units of  $k_B T$  as a function of Cc CM. This panel is taken from ref. (Northrup et al. 1988) with permission from Science. **d** The blue and green meshes indicate reweighted atomic probability density maps for the overall distribution of Cc molecules obtained from, respectively, Monte-Carlo simulations (same as in **b**) and PRE-based ensemble calculations (same as in Fig. 5a), both plotted at a threshold of 20% maximum. The dominant form of the complex is in the same orientation as in Fig. 1a. In the bottom panel, SL atoms are removed for clarity

#### Practical considerations for the analysis of protein encounters by PRE NMR spectroscopy

As shown here, a good experimental coverage of the entire conformational space available to the interacting proteins is essential for an accurate description of the encounter state. In our case, the experimental sampling was achieved by

varying the conjugation site of the paramagnetic probe on the surface of CcP. We estimate that 13–20 uniformly spaced SL attachment positions, located outside the crystallographic binding site, are required to provide an adequate PRE coverage for Cc–CcP complex at  $p = 0.1$ – $0.01$  (Table 1). This means that, in addition to our dataset, at least 3–10 extra SLs would be needed to complete the encounter map. (In practice, this number is expected to be higher due to non-uniform distribution of the already introduced SLs.)

To transpose our findings to other systems involving globular proteins, we note that attachment of one SL per each 190–300  $\text{\AA}^2$  of the total surface area (SA)—or 160–250  $\text{\AA}^2$  of the SA excluding the binding site—is required for the complete coverage at  $p = 0.01$ – $0.1$  (Table 1). In other words, a SL should be placed every 5–10 protein surface residues (defined as those with solvent-accessible SA  $> 10 \text{\AA}^2$ ). The introduced SL must not perturb the biomolecular interaction studied (e.g. steric clashes with the dominant binding form or substitutions of charged residues altering electrostatic potentials should be avoided), which limits the choice of the SL attachment locations. In practical terms, introduction of a SL at each site necessitates preparation of the corresponding single-cysteine protein variant. Thus, a comprehensive, SL-based PRE NMR encounter mapping requires a significant experimental effort, approaching that of labor-intensive EPR studies (Crane et al. 2005, 2006, 2008).

Several strategies can be followed to expedite the analysis. First, if possible, paramagnetic labeling of the smaller protein should be carried out as it requires fewer conjugation sites for a good experimental coverage. In case of Cc–CcP, 5–7 SLs attached to Cc (i.e. approximately one-third of those needed for CcP; Table 1) would be enough for the complete coverage at  $p = 0.1$ – $0.01$ . Further, paramagnetic tagging of both interacting proteins, one at a time, would allow one to decrease the number of attachment sites even more. The main limiting factors of this approach are the quality of the NMR spectra and the availability of backbone assignments for the bigger protein, which have thwarted its application to the Cc–CcP complex. Second, the use of stronger paramagnetic labels (e.g. an EDTA–Mn chelate or lanthanide-containing probes; Clore and Iwahara 2009; Su and Otting 2010) would significantly decrease the number of conjugation sites required for a good experimental coverage and could offer a number of additional advantages. For example, pseudo-contact shifts originating from the introduced lanthanide atoms could provide an independent means of verifying the structure of the dominant binding form in solution, and the use of rigid, two-armed, paramagnetic tags (Keizers et al. 2007, 2008) would obviate the need for extensive ensemble

averaging of the measured PREs, thus allowing for a more accurate description of the encounter state.

## Conclusions

The spatiotemporal mapping approach presented here provides a reliable estimate of the experimental coverage and, at higher coverage levels, allows to delineate the conformational space sampled by the minor species. As shown in recent studies of Cc–CcP (Bashir et al. 2010) and other complexes formed by charged proteins (Kim et al. 2008), electrostatics-based MC simulations afford a robust estimate of the encounter state population, which is further confirmed by ensemble refinement performed in this work. However, to obtain an accurate description of the encounter state's conformational space, further refinement of MC solutions appears to be necessary. The combination of methods employed here for the analysis of Cc–CcP encounter state illustrates a general approach for comprehensive visualization of transient species in biomolecular systems.

**Acknowledgments** A.N.V. is an FWO Post-Doctoral Researcher. M. U. acknowledges the funding by the Netherlands Organisation for Scientific Research (NWO), VICI grant, no. 700.58.441. We acknowledge the access to the VUB/ULB Computing Centre and thank Lieven Butts for the help with debugging the scripts and G. Marius Clore for critical reading of the manuscript.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Adam G, Delbruck M (1968) Structural chemistry and molecular biology. Freeman, San Francisco, pp 198–215
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98:10037–10041
- Bashir Q, Volkov AN, Ullmann GM, Ubbink M (2010) Visualization of the encounter ensemble of the transient electron transfer complex of cytochrome *c* and cytochrome *c* peroxidase. *J Am Chem Soc* 132:241–247
- Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* 39:5355–5365
- Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J (2007) Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. *J Am Chem Soc* 129:12786–12794
- Bertini I, Giachetti A, Luchinat C, Parigi G, Petoukhov MV, Pierattelli R, Ravera E, Svergun DI (2010) Conformational space of flexible biological macromolecules from average data. *J Am Chem Soc* 132:13553–13558
- Brünger AT, Clore GM, Gronenborn AM, Saffrich R, Nilges M (1993) Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* 261:328–331
- Chance B, Devault D, Legallais V, Mela L, Yonetani T (1967) Kinetics of electron transfer reactions in biological systems. In: Claesson S (ed) Fast reactions and primary processes in chemical reactions. Interscience, New York, pp 437–464
- Clore GM (2008) Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement. *Mol Biosyst* 4:1058–1069
- Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* 109:4108–4139
- Cooper DB, Smith VF, Crane JM, Roth HC, Lilly AA, Randall LL (2008) SecA, the motor of the secretion machine, binds diverse partners on one interactive surface. *J Mol Biol* 382:74–87
- Crane JM, Mao C, Lilly AA, Smith VF, Suo Y, Hubbell WL, Randall LL (2005) Mapping of the docking of SecA onto the chaperone SecB by site-directed spin labeling: insight into the mechanism of ligand transfer during protein export. *J Mol Biol* 353:295–307
- Crane JM, Suo Y, Lilly AA, Hubbell WL, Randall LL (2006) Sites of interaction of a precursor on the export chaperone SecB mapped by site-directed spin labeling. *J Mol Biol* 363:63–74
- DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto
- Gabdoulline RR, Wade RC (2001) Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations. *J Mol Biol* 306:1139–1155
- Gardner RJ, Longinetti M, Sgheri L (2005) Reconstruction of orientations of a moving protein domain from paramagnetic data. *Inv Probl* 21:879–898
- Hartl MJ, Schweimer K, Reger MH, Schwarzingger S, Bodem J, Rösch P, Wöhrl BM (2010) Formation of transient dimers by a retroviral protease. *Biochem J* 427:197–203
- Iwahara J, Clore GM (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* 440:1227–1230
- Iwahara J, Anderson DE, Murphy EC, Clore GM (2003) EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. *J Am Chem Soc* 125:6634–6635
- Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against  $^1\text{H}$  paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 126:5879–5896
- Iwahara J, Tang C, Clore GM (2007) Practical aspects of  $^1\text{H}$  transverse paramagnetic relaxation enhancement measurements on macromolecules. *J Magn Reson* 184:185–195
- Keizers PHJ, Desreux JF, Overhand M, Ubbink M (2007) Increased paramagnetic effect of a lanthanide protein probe by two-point attachment. *J Am Chem Soc* 129:9292–9293
- Keizers PHJ, Saragliadis A, Hiruma Y, Overhand M, Ubbink M (2008) Design, synthesis, and evaluation of a lanthanide chelating protein probe: CLaNP-5 yields predictable paramagnetic effects independent of environment. *J Am Chem Soc* 130:14802–14812
- Kim YC, Tang C, Clore GM, Hummer G (2008) Replica exchange simulations of transient encounter complexes in protein-protein association. *Proc Natl Acad Sci U S A* 105:12855–12860
- Longinetti M, Luchinat C, Parigi G, Sgheri L (2006) Efficient determination of the most favoured orientations of protein domains from paramagnetic NMR data. *Inv Probl* 22:1485–1502

- Northrup SH, Boles JO, Reynolds JCL (1988) Brownian dynamics of cytochrome *c* and cytochrome *c* peroxidase association. *Science* 241:67–70
- Pelletier H, Kraut J (1992) Crystal structure of a complex between electron transfer partners, cytochrome *c* peroxidase and cytochrome *c*. *Science* 258:1748–1755
- Schreiber G, Haran G, Zhou H-X (2009) Fundamental aspects of protein-protein association kinetics. *Chem Rev* 109:839–860
- Schwieters CD, Clore GM (2002) Reweighted atomic densities to represent ensembles of NMR structures. *J Biomol NMR* 23: 221–225
- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:66–74
- Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc* 48:47–62
- Solomon I (1955) Relaxation processes in a system of two spins. *Phys Rev* 99:559–565
- Solomon I, Bloembergen N (1956) Nuclear magnetic interactions in the HF molecule. *J Chem Phys* 25:261–266
- Su X-C, Otting G (2010) Paramagnetic labelling of proteins and oligonucleotides for NMR. *J Biomol NMR* 46:101–112
- Tang C, Iwahara J, Clore GM (2006) Visualization of transient encounter complexes in protein-protein association. *Nature* 444:383–386
- Tang C, Ghirlardo R, Clore GM (2008a) Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J Am Chem Soc* 130:4048–4056
- Tang C, Louis JM, Aniana A, Suh J-Y, Clore GM (2008b) Visualizing transient events in amino-terminal autoproteolysis of HIV-1 protease. *Nature* 455:693–696
- Ubbink M (2009) The courtship of proteins: understanding the encounter complex. *FEBS Lett* 583:1060–1066
- Volkov AN, Worrall JAR, Holtzmann E, Ubbink M (2006) Solution structure and dynamics of the complex between cytochrome *c* and cytochrome *c* peroxidase determined by paramagnetic NMR. *Proc Natl Acad Sci U S A* 103:18945–18950
- Volkov AN, Bashir Q, Worrall JAR, Ullmann GM, Ubbink M (2010) Shifting the equilibrium between the encounter state and the specific form of a protein complex by interfacial point mutations. *J Am Chem Soc* 132:11487–11495
- Worrall JAR, Liu Y, Crowley PB, Nocek JM, Hoffman BM, Ubbink M (2002) Myoglobin and cytochrome *b*<sub>5</sub>: a nuclear magnetic resonance study of a highly dynamic protein complex. *Biochemistry* 41:11721–11730
- Xu X, Reinle W, Hannemann F, Konarev PV, Svergun DI, Bernhardt R, Ubbink M (2008) Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. *J Am Chem Soc* 130:6395–6403
- Young LJ, Caughey WS (1987) Autoreduction phenomena of bovine heart cytochrome *c* oxidase and other metalloproteins. *J Biol Chem* 262:15019–15025