Dynamics in electron transfer protein complexes

Bashir, Q.

Chapter 5

Shifting the equilibrium between the encounter state and the specific form of a protein complex by interfacial point mutations

This work has been published as part of

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, 11487-95.

Abstract

Recent experimental studies have confirmed a long-held view that protein complex formation proceeds via a short-lived encounter state. The population of this transient intermediate, stabilized mainly by long-range electrostatic interactions, varies among different complexes. Here we show that the occupancy of the encounter state can be modulated across a broad range by single point mutations of interfacial residues. Using a combination of Monte Carlo simulations and paramagnetic relaxation enhancement NMR spectroscopy, we illustrate that it is possible to both enhance and diminish the binding specificity in an electron transfer complex of yeast cytochrome *c* (Cc) and cytochrome *c* peroxidase. The Cc T12A mutation decreases the population of the encounter to 10 % as compared with 30 % in the wild-type complex. More dramatically, the Cc R13A substitution reverses the relative occupancies of the stereospecific and the encounter forms, with the latter now being the dominant species with the population of 80 %. This finding indicates that the encounter state can make a large contribution to the stability of a protein complex. Also, it appears that by adjusting the amount of the encounter through a judicious choice of point mutations, we can remodel the energy landscape of a protein complex and tune its binding specificity.

Introduction

Cellular function depends critically on the ability to transmit, receive, and integrate various signals arising within and outside the cell. In order to propagate a biochemical event such as phosphorylation or electron transfer (ET) to distant cellular sites fast and efficiently, protein complexes mediating relay processes are optimized for high turnover. As a consequence, such transient interactions are short-lived and typically have high association (k_{on}) and dissociation (k_{off}) rate constants, resulting in weak binding ($K_D = k_{off}/k_{on}$ in μM - mM range)[9,14]. Another constraint is imposed by the necessity to recognize multiple binding partners, which limits the specificity. Thus, transient complexes have evolved to function within narrow margins of affinity and specificity in order to maximize productive binding and minimize wasteful interactions with other components of the cellular milieu.

A good example of a biomolecular transient system is the ET protein complex of yeast cytochrome *c* (Cc) and cytochrome *c* peroxidase (CcP). The proteins are localized in 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[180]. The catalytic cycle of CcP begins with the coordination of peroxide to the ferric haem. Then, bound peroxide undergoes rapid heterolytic cleavage, producing a molecule of water and an intermediate referred to as compound ES[82]. In this state, the haem iron is oxidized to an oxyferryl Fe (IV) group and an active-site tryptophan residue (W191) is oxidized to a cationic indole radical[83-87]. Then, in two successive one-electron transfer reactions, Cc reduces compound ES back to resting-state CcP[88,89].

The stoichiometry of Cc - CcP complex has been a subject of debate. The crystal[81] and solution[19] structures of the Cc - CcP complex show a single 1 : 1 binding site. However, several studies proposed that the interaction of Cc and CcP occurs with a 2 : 1 stoichiometry, involving non-overlapping binding domains[90-93,97-99]. Yet other studies led to the conclusion that only a single binding site exists[94-96], and recent mutagenesis and cross-linking work by Erman and coworkers strongly supports this conclusion[181-183]. Paramagnetic NMR experiments revealed a significant degree of dynamics within the Cc - CcP complex[19] and established that the encounter state, which precedes the

formation of the specific binding geometry[17], represents 30 % of the complex in solution (chapter 4)[184]. The encounter state, believed to accelerate the search of the optimal binding geometry, is best represented by an ensemble of binary complexes, with the proteins assuming many different orientations. Recently, we have shown that this ensemble can be described well assuming only electrostatic interactions (chapter 4)[184], in agreement with earlier theoretical docking simulations[102,103].

The delicate balance between the encounter state and the specific complex may well be functionally relevant. An electrostatically-driven encounter enhances the association rate, but exhibits low ET activity due to the large distance between the redox centers. The specific complex brings these centers within the range of rapid ET (chapter 4)[184], yet it must not be very stable to avoid a slow turn-over. Thus, we wondered whether the balance between an encounter and the specific complex can be modulated. Here we show that it is possible to both decrease and increase the encounter population by single point mutations of Cc residues in the binding interface, which suggests a subtle yet efficient way to remodel the energy landscape of transient protein-protein interactions.

Our prior work has established the existence of minor species in Cc-CcP complex[19] and laid methodological foundations for PRE NMR analysis of transient encounter states in protein-protein interactions (chapter 4)[184], whereas this work focuses entirely on the study of the equilibrium between the encounter state and the specific form using a panel of interfacial mutants. Armed with a variety of biophysical and computational tools, one of which is PRE NMR, we address the delicate balance between the specific form and the encounter state and assess the contribution of the latter to the total binding energy of the complex. A vast body of literature investigating the impact of interfacial mutations on protein-protein association reactions addresses the equilibrium between the free and bound forms, but not the transition between the encounter state and the specific binding geometry. To the best of our knowledge, the latter equilibrium, let alone its perturbation by surface mutations, has never been quantified experimentally, which is the focus of this study.

Methods

Sample preparation

Both unlabelled and U-^{15}N labelled T-5A/C102T Cc, referred to as wt Cc, were expressed in *Escherichia coli* and purified as described before[112,144]. Cc mutations were introduced by site-directed mutagenesis using the Quik Change polymerase chain reaction protocol (Stratagene) with the plasmid encoding the gene for wt Cc as a template. The constructs were verified by DNA sequencing, and the proteins produced and purified analogously to wt Cc. Wild-type CcP and its single-cysteine variants used in the PRE analysis were isolated from *E. coli* following the published procedures[19,145]. CcP mutants were conjugated, one at a time, with a paramagnetic spin-radical [MTSL, (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] as described[19]. The yield of labelling, determined by EPR or a spectrophotometric dithiodipyridine assay[166], was close to 100 %. Both Cc and CcP were in the ferric, oxidized form in all experiments.

NMR spectroscopy

PRE NMR samples contained 0.3 - 0.4 mM wt or mutant ^{15}N Cc and one molar equivalent of a single-cysteine CcP variant labelled with a paramagnetic spin-radical (MTSL) or its diamagnetic analogue (MTS). For chemical shift perturbation analysis, samples containing 0.5 mM of ^{15}N Cc mutants or their 1:1 complexes with wt CcP were used. Measurements were performed in 20 mM sodium phosphate, 0.1 M NaCl pH 6.0, 6 % D_2O for lock, and 0.1 mM each of TSP [3-(trimethylsilyl)-propionic acid-D4, sodium salt] and $CH_3CO^{15}NH_2$ as internal references at 303 K on a Bruker DMX-600 spectrometer equipped with a cryogenic probe. The pH of the sample was checked before each titration point and, if necessary, adjusted to 6.00 ± 0.05 with concentrated NaOH or HCl solutions. Two-dimensional [^{15}N, ^{1}H] HSQC spectra were acquired with 1024 and 128 complex points in the direct and indirect dimensions, respectively. All data were processed with Azara 2.7 (provided by W. Boucher and Department of

Biochemistry, University of Cambridge; available from http://www.bio.cam.ac.uk/azara/) and analysed in Ansig for Windows[115]. Assignments of the ^{15}N and ^{1}H nuclei of wt Cc were taken from previous work[101]. The averaged amide chemical shift perturbations ($\Delta\delta_{avg}$) were derived from $\Delta\delta_{avg} = [(\Delta\delta_N)^2/50 + (\Delta\delta_H)^2/2]^{1/2}$, where $\Delta\delta_N$ and $\Delta\delta_H$ are the differences in the chemical shifts between free and bound form for the amide nitrogen and proton, respectively. In order to obtain $\Delta\delta_{avg}$ extrapolated to 100 % bound form, the original $\Delta\delta_{avg}$ was divided by the molar fraction of Cc bound, which, at the protein concentrations used, was 0.9 (wt), 0.97 (T12A), 0.82 (R13K), or 0.57 (R13A). Binding constants from Table 5.1 were used to calculate the amount of Cc bound.

PRE measurements

PREs were calculated from peak intensities in the HSQC spectra of paramagnetic (I_{para}) and diamagnetic control (I_{dia}) samples as

$$\frac{I_{para}}{I_{dia}} = \frac{R_2 \exp(-t\Gamma_2)}{R_2 + \Gamma_2} \qquad (5.1)$$

Where R_2 is the transverse relaxation rate of Cc amide protons in the diamagnetic complex; Γ_2 is the PRE and t is the total INEPT evolution time of the HSQC[168]. To account for minor differences in protein concentrations between the two samples, the obtained I_{para}/I_{dia} were normalized as described before[184]. For the resonances that disappear in the paramagnetic spectrum, an upper limit for I_{para} was estimated from the standard deviation of the noise level of the spectrum. For each peak, R_2 was estimated from the width at half-height ($\Delta\upsilon_{1/2}$) of a Lorentzian fit in the proton dimension by using $R_2 = \pi\Delta\upsilon_{1/2}$.

Monte Carlo simulations

A Boltzmann ensemble of encounter complex geometries was generated by Metropolis Monte Carlo simulations[169,184]. Protein coordinates were taken from the crystal structure

of the complex (PDB entry 2PCC[81]), and protein molecules kept rigid during the simulations. Cc was allowed to move in the electrostatic potential of CcP, and moves were accepted or rejected according to the Metropolis criterion. The energy of each protein-protein orientation was calculated by multiplying the charges of Cc with the electrostatic potential of CcP. An exclusion grid was used to prevent steric overlap between the proteins. When the centers of mass of the two molecules were within 40 Å, the protein coordinates were saved every 1000 steps. In total, 1500 to 2000 configurations were retained for the analysis. In this approach, protein-protein orientations with large intermolecular distances contribute negligibly to the free energy change of protein complex formation and will also be invisible in the PRE analysis.

Mutations were introduced *in silico*, followed by the addition of hydrogen atoms and energy minimization in Xplor-NIH[150]. The simulation temperature and the ionic strength were set to 303 K and 0.12 M, respectively, to match the experimental conditions. The Poisson-Boltzmann electrostatic potential was calculated by the MEAD program suite[170] using atomic partial charges of the Charmm force field[171]. The dielectric constants used for the protein and water were 4 and 80, respectively.

Analysis of the encounter state

For each CcP conjugation site, the orientation of the attached SL was randomised in Xplor-NIH and an ensemble of four representative conformers used for subsequent data analysis. For every structure of the simulated encounter ensemble, distances between Cc backbone amide protons and the oxygen atom in each of the four SL conformers were measured and averaged as described before[184]. The effective distances, r_i, were used to calculate PREs for Cc atoms in the encounter state ($\Gamma_{2,i}^{encounter}$):

$$\Gamma_i^{encounter} = \frac{\gamma^2 g^2 \beta^2}{20 r_i^6}\left(4\tau_c + \frac{3\tau_c}{1+\omega_h^2\tau_h^2}\right) \qquad (5.2)$$

where τ_c is the correlation time of the electron-nucleus vector (16 ns)[19]; ω_h and γ are proton Larmor frequency and gyromagnetic ratio, respectively; g is the electronic g-

factor; and β is the Bohr magneton. The observed PRE is the sum of the population-weighted contributions of the specific form and the encounter state:

$$\Gamma_{2,i}^{obs} = p\Gamma_{2,i}^{encounter} + (1-p)\Gamma_{2,i}^{specific} \qquad (5.3)$$

The $\Gamma_{2,i}^{encounter}$ and $\Gamma_{2,i}^{specific}$ were back-calculated from the simulated encounter state and the crystal structure of the complex, respectively, following a published protocol using a four-conformer ensemble representation of the attached SL to account for its flexibility (chapter 4)[19,35,184]. To obtain the population of the encounter state (p) in a given complex, the agreement between the observed and the calculated PREs at varying p values was estimated using the Q-factor[184] (eq 5.4), and p values at lowest Q were selected.

$$Q = \sqrt{\sum_i \left(\Gamma_{2,i}^{obs} - \Gamma_{2,i}^{calc}\right)^2 / \sum_i \left(\Gamma_{2,i}^{obs} + \Gamma_{2,i}^{calc}\right)^2} \qquad (5.4)$$

The total calculated PRE ($\Gamma_{2,i}^{calc}$) is computed as a sum of the population-weighted $\Gamma_{2,i}^{encounter}$ and $\Gamma_{2,i}^{specific}$ (eq 5.3), with the latter back-calculated from the crystal structure of the complex. Finally, an agreement between observed and calculated PREs at each *p* value is quantified (eq 5.4). The uncertainties in the Q factors were obtained by standard propagation of the error in $\Gamma_{2,i}^{obs}$, derived from the experimental uncertainties in I_{para} and I_{dia} (chapter 4)[184]. For all Cc mutants, only the data from SLs that produced intermolecular paramagnetic effects (N38C, N200C, and T288C) were included in the analysis. As judged from the calculated Q factors, control MC runs employing proteins with randomised side-chain orientations or different sets of SL conformers produced results very similar to those presently described for all mutant complexes (data not shown). All structures were visualized with PyMOL (http://www.pymol.org).

Theory

Decomposition of the free binding energy into the contributions of the specific and encounter forms

The formulae below were derived by Prof. M. Ubbink and Dr. A. Volkov with help from Dr. E. Blokhuis (Leiden University) and Dr. I Lasters (Algonomics NV, Belgium). Cc and CcP form a complex <Cc:CcP> with a degenerate energy level consisting of two sub-levels: those of a specific form Cc:CcP and an encounter state $(Cc:CcP)^*$. Complex formation can be described by the scheme 5.5:

$$Cc + CcP \rightleftharpoons \langle Cc:CcP \rangle \qquad (5.5)$$

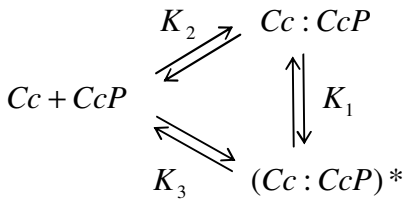
with the equilibrium constant (K_0) given by

$$K_0 = \frac{a(\langle Cc:CcP \rangle)}{a(Cc)a(CcP)} = \frac{[\langle Cc:CcP \rangle]}{[Cc][CcP]} \times \frac{\gamma(\langle Cc:CcP \rangle)}{\gamma(Cc)\gamma(CcP)} \approx \frac{[\langle Cc:CcP \rangle]}{[Cc][CcP]} \qquad (5.6)$$

where a and γ are activities and activity coefficients, respectively. In dilute protein solutions used in this study, the latter approach 1 M^{-1}. Note that K_0 and other equilibrium constants defined below are dimensionless entities. The populations of the free proteins (P_0) and the <Cc:CcP> complex (P_1) in scheme 5.5 are:

$$P_0 = \frac{1}{1+2K_0} \qquad P_1 = \frac{2K_0}{1+2K_0} \qquad (5.7)$$

where the factor 2 in $2K_0$ term reflects the two-fold degeneracy of the energy level.
Formation of the specific form and the encounter state can be described by the scheme 5.8:



(5.8)

where the corresponding equilibrium constants and populations of the free proteins (P_0), the specific form (P_2), and the encounter state (P_3) are given by

$$K_1 = \frac{[Cc:CcP]}{[(Cc:CcP)^*]} \qquad K_2 = \frac{[Cc:CcP]}{[Cc][CcP]} \qquad K_3 = \frac{[(Cc:CcP)^*]}{[Cc][CcP]} \tag{5.9}$$

$$P_0 = \frac{1}{1+K_2+K_3} \qquad P_2 = \frac{K_2}{1+K_2+K_3} \qquad P_3 = \frac{K_3}{1+K_2+K_3} \tag{5.10}$$

Equating P_0 terms from (eq 5.7) and (eq 5.10), and noting that $K_1 = K_2/K_3$, yields $K_2 = 2K_0K_1/(1 + K_1)$ and $K_3 = 2K_0/(1 + K_1)$. Finally, using the experimental free binding energy for <Cc:CcP> (ΔG_0) and the population of the encounter state (p) derived from PRE NMR and given by eq 5.11

$$p = \frac{[(Cc:CcP)^*]}{[Cc:CcP]+[(Cc:CcP)^*]} = \frac{1}{K_1+1} \tag{5.11}$$

allows to derive the free energy contributions of the specific form (ΔG_2) and the encounter state (ΔG_3):

$$\Delta G_2 = -RT\ln(K_2) = -RT\ln\left(\frac{2K_0K_1}{1+K_1}\right) = \Delta G_0 + RT\ln\left(\frac{1+K_1}{2K_1}\right) = \Delta G_0 - RT\ln[2(1-p)] \tag{5.12}$$

$$\Delta G_3 = -RT\ln(K_3) = -RT\ln\left(\frac{2K_0}{1+K_1}\right) = \Delta G_0 + RT\ln\left(\frac{1+K_1}{2}\right) = \Delta G_0 - RT\ln(2p) \tag{5.13}$$

The errors in ΔG_2 and ΔG_3 are propagated from the uncertainties in ΔG_0 and p as

$$\delta(\Delta G_2) = \sqrt{\delta^2(\Delta G_0) + [RT\delta(p)/(1-p)]^2} \text{ and}$$

$$\delta(\Delta G_3) = \sqrt{\delta^2(\Delta G_0) + (RT\delta(p)/p)^2} \qquad (5.14)$$

Results and Discussion

Characterisation of variant Cc - CcP complexes

The structure of the specific Cc - CcP complex is shown in Figure 5.1A, with the binding surface of Cc detailed in Figure 5.1B. Cc residues T12 and R13, located in the center of the interface, mediate interactions with a variety of reaction partners[29,124,159], and the latter is part of a binding hotspot in the complex with CcP[141]. Given the importance of these residues in macromolecular recognition, we sought to investigate their role in the specificity of Cc - CcP interaction. To this end, alanine-substituted (T12A and R13A) and charge-conserved (R13K) Cc mutants were prepared. To ascertain that the introduced mutations do not perturb the structure of the protein, we compared the NMR chemical shifts of Cc variants with those of the wt protein (Figure 5.2). Except for several residues adjacent to the mutation sites, the chemical shifts of the variants are identical to those of wt Cc, indicating that the mutations do not perturb the global fold of the protein. R13K Cc exhibits the smallest $\Delta\delta$ (Figure 5.2C), which is expected given the similarity of lysine and arginine side-chains.

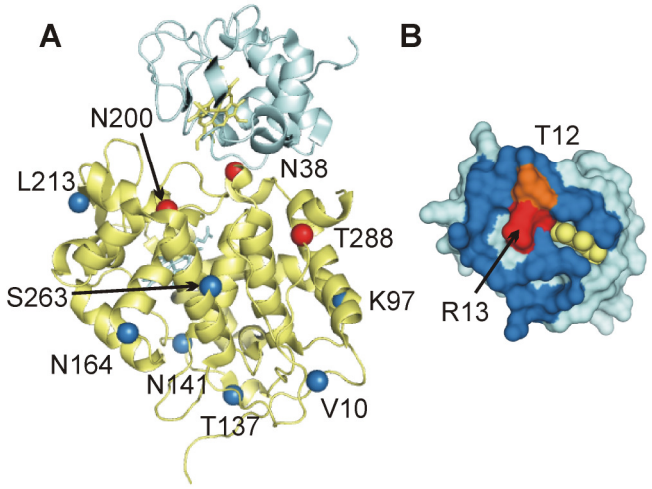


Figure 5.1. Specific form of Cc - CcP complex. (A) Crystallographic Cc - CcP orientation. Cc and CcP are in blue and yellow, with haem groups in sticks. C_α atoms of CcP residues used for spin-labelling are shown as spheres coloured according to whether the attached SL exhibits intermolecular PREs (red) or not (blue). (B) Binding interface of Cc in the specific complex. The residues that loose solvent accessible surface area upon complex formation with CcP are dark-coloured; the haem group is spacefilled and shown in yellow; residues T12 and R13 are coloured orange and red, respectively. The figure was drawn from the crystal structure of the complex (PDB entry 2PCC[81]).

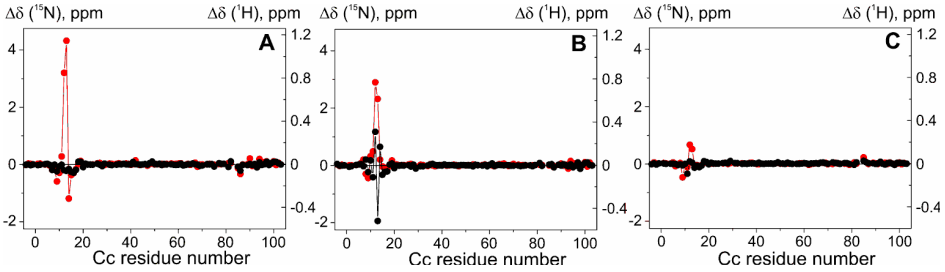


Figure 5.2. Effects of mutations on the chemical shifts of Cc backbone amides. Changes in the chemical shifts (Δδ) of backbone amide nitrogens (red circles) and protons (black circles) for (A) R13A, (B) T12A, and (C) R13K Cc, relative to those of wt Cc.

The binding constants (K_B) of the wt and mutant complexes were determined by NMR, or taken from literature (Table 5.1 and Figure 5.3). In all cases, titration curves fit well

to a 1 : 1 stoichiometry model, suggesting that no binding of a second Cc molecule takes place under our experimental conditions. As can be seen from Table 5.1, the Cc R13K substitution leads to a 4-fold decrease in K_B, whereas the R13A mutation reduces it 100-fold. Interestingly, a 10-fold increase in K_B is observed in T12A Cc - CcP complex.

Table 5.1. Binding constants of Cc - CcP complexes.

Cc	K_B, 10^5 M^{-1}
wt	1.9 ± 0.1[a]
T12A	17 ± 1[b]
R13K	0.54 ± 0.03
R13A	0.06 ± 0.002[a]

All values were determined in 20 mM sodium phosphate 0.1 M NaCl pH 6.0 at 303 K.

[a] Values taken from ref[141] (chapter 3).

[b] Values taken from ref[185]

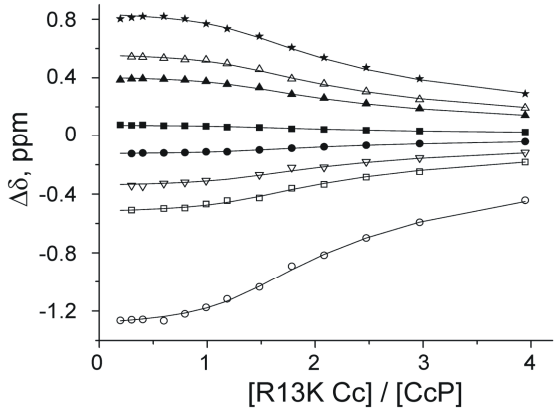


Figure 5.3. Binding constants. Chemical shift perturbations (Δδ) of ^{15}N R13K Cc backbone amide protons (open symbols) and nitrogen atoms (filled symbols) observed in a series of 2D [^{1}H, ^{15}N] HSQC spectra for residues L9 (squares), T12 (circles), C14 (inverted triangles), and Q16 (triangles) upon titration into wt CcP. Also plotted is Δδ for the Cc haem 3-methyl protons (stars), followed in 1D ^{1}H spectra. The curves represent the best fit to a 1:1 binding model[141] with the shared binding constant of $(5.4 \pm 0.3) \times 10^4$ M^{-1}.

Chemical shift perturbations of Cc residues in variant and wt complexes with CcP exhibit similar patterns, indicating that mutant and wt Cc proteins employ the same interface to bind CcP (Figure 5.4). However, the magnitude of the binding shifts (extrapolated to 100 % Cc bound, see Methods) differs, decreasing in the order T12A > wt > R13K > R13A Cc. The absolute size of chemical shift perturbations has been reported to correlate with the dynamics in transient protein complexes, with smaller shifts indicating increased mobility[26,29,124]. Thus, it appears that R13A and R13K Cc - CcP are more dynamic than the wt complex, whereas the protein mobility in T12A Cc - CcP is nearly unaffected or even slightly decreased.

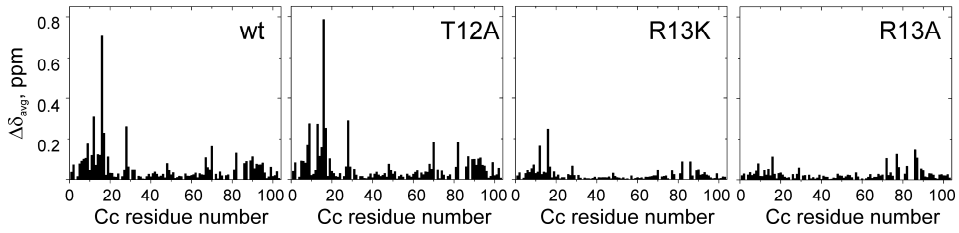


Figure 5.4. NMR chemical shift perturbations ($\Delta\delta_{avg}$) for the ^{15}N Cc - CcP complexes studied in this work. All $\Delta\delta_{avg}$ are extrapolated to 100% bound form (see Methods), and thus do not reflect differences in the binding constants.

Intermolecular PREs

In order to study the variant Cc - CcP complexes further, ten single-cysteine CcP mutants were produced and conjugated with a paramagnetic spin-label (SL) (Figure 5.1A). The unpaired electron of CcP-SL causes an intermolecular PRE of the nearby Cc nuclei, which depends on the distance between the unpaired electron and the observed nucleus and is readily quantified by NMR. For every Cc variant, multiple PRE profiles were measured in binary complexes with different CcP-SL (Figures 5.5 and 5.5-5.8). The protein interactions studied here are in fast exchange on the NMR time scale; thus, the linewidth of Cc resonances is a weighted average of those in the free (12 kDa) and CcP-bound (46 kDa) forms. At the protein concentrations used, the fraction of bound Cc decreases in the order of T12A (96 ± 1 %) > R13K (79 ± 2 %) > R13A (50 ± 3 %), which explains higher signal-to-noise ratios for weaker R13-

substituted complexes (Figures 5.5 and 5.6-5.8). Varying noise levels for CcP-SL complexes with the same Cc mutant are due to different sample concentrations, total acquisition time, or both. Only five PRE profiles were obtained for T12A Cc. As T12A Cc - CcP data are very similar to those of the wt complex, the datasets with SL positions V10C, K97C, N141C, N164C, and L213C, which showed no effect in the wt complex (chapter 4)[184], were not collected.

For all Cc mutants, only three SLs attached close to the crystallographic binding site (positions N38C, N200C and T288C, coloured red in Figure 5.1A) give rise to observable PREs, whereas others show no effects (Figures 5.5 and 5.5-5.8). Overall, the distribution of the PREs is highly similar to that of wt Cc – CcP (chapter 4)[184], suggesting that the introduced Cc mutations do not significantly alter the conformational space sampled by the proteins within the complex.

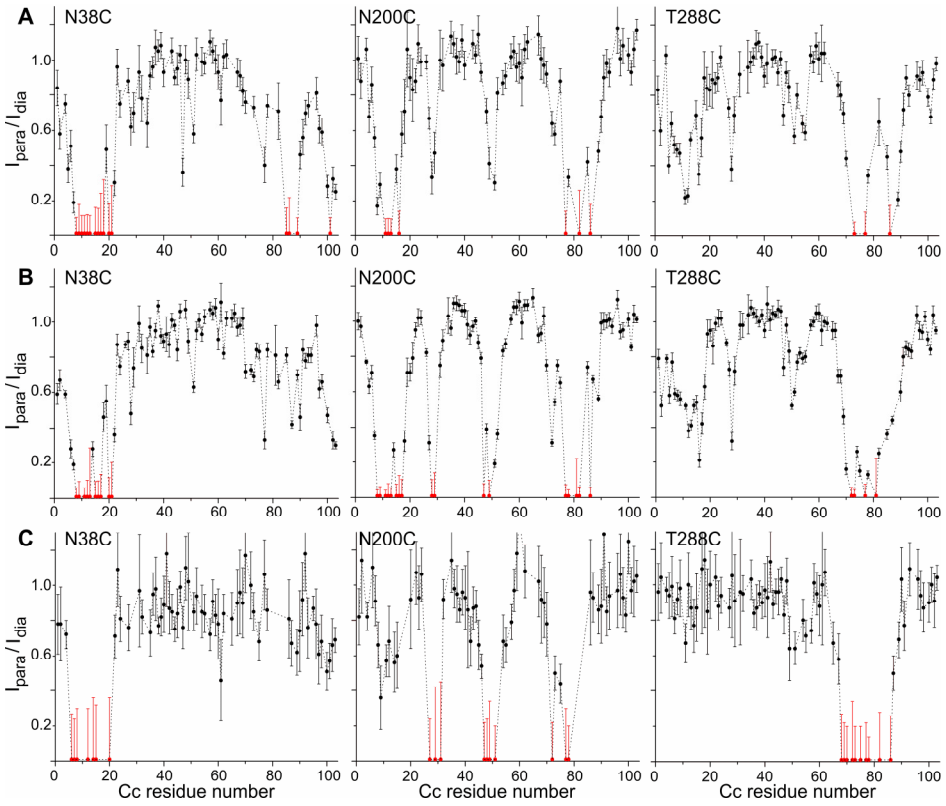


Figure 5.5. Paramagnetic effects in Cc - CcP-SL complexes. Relative [^{1}H, ^{15}N] HSQC intensities of (A) R13A, (B) R13K, and (C) T12A Cc backbone amides in the complex with CcP labeled with a paramagnetic SL (I_{para}) or a diamagnetic analogue (I_{dia}) at positions N38C, N200C and T288C. For the residues whose resonances disappear in the paramagnetic spectrum, the upper limit of I_{para}/I_{dia} was estimated from the noise level. The error bars denote standard deviations, derived from spectral noise levels using standard error propagation procedures. Profiles for the other SL positions, showing no measurable effects, are shown in Figures 5.6-5.8.

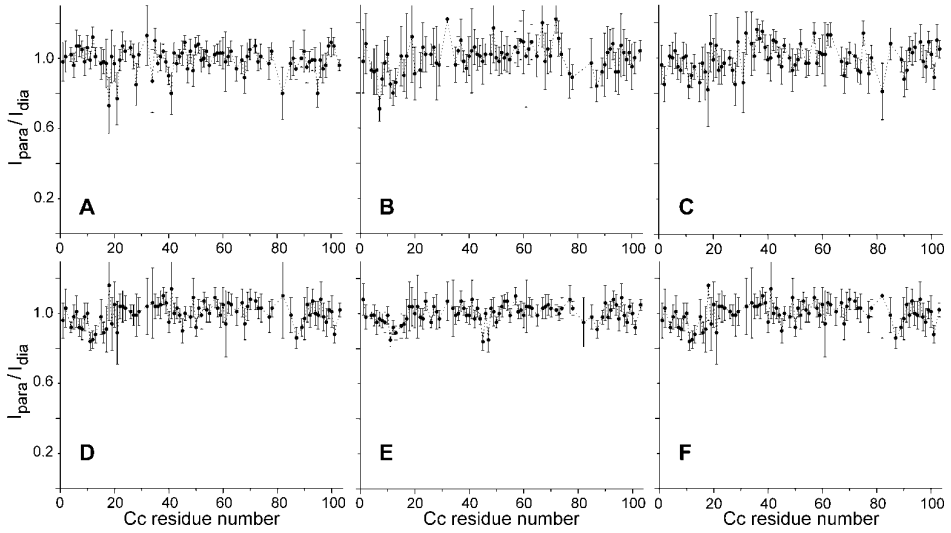


Figure 5.6. Paramagnetic effects in R13A Cc - CcP-SL complexes. Relative [^{1}H, ^{15}N] HSQC intensities of Cc backbone amides in the complex with CcP labeled with a paramagnetic SL (I_{para}) or a diamagnetic analogue (I_{dia}) at positions (A) V10C, (B) K97C, (C) T137C, (D) N141C, (E) N164C, and (F) L213C. The error bars denote standard deviations, derived from spectral noise levels using standard error propagation procedures.

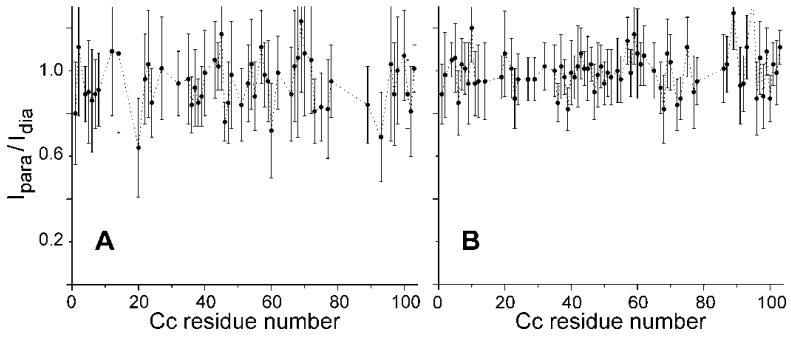


Figure 5.7. Paramagnetic effects in T12A Cc - CcP-SL complexes. Relative [^{1}H, ^{15}N] HSQC intensities of Cc backbone amides in the complex with CcP labeled with a paramagnetic SL (I_{para}) or a diamagnetic analogue (I_{dia}) at positions (A) T137C, and (B) S263C. The error bars denote standard deviations, derived from spectral noise levels using standard error propagation procedures.

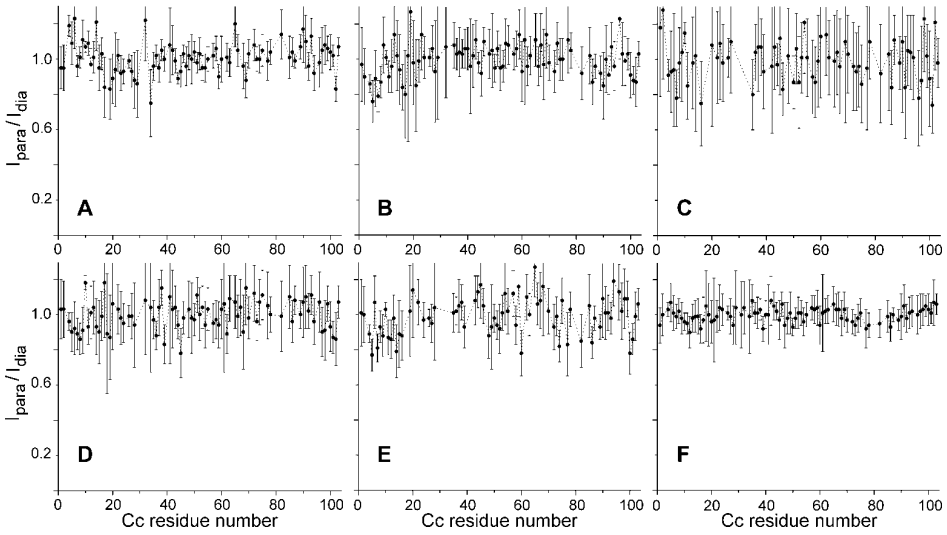


Figure 5.8. Paramagnetic effects in R13K Cc - CcP-SL complexes. Relative [^{1}H, ^{15}N] HSQC intensities of Cc backbone amides in the complex with CcP labeled with a paramagnetic SL (I_{para}) or a diamagnetic analogue (I_{dia}) at positions (A) V10C, (B) K97C, (C) N141C, (D) N164C, (E) L213C, and (F) S263C. The error bars denote standard deviations, derived from spectral noise levels using standard error propagation procedures.

Encounter ensemble simulations

The measured PRE (Γ_2^{obs}) is a population-weighted average of the contributions from all protein-protein orientations (eq 5.2) and, as such, contains the information on both the specific form and the encounter state[22]. The PRE for the former ($\Gamma_2^{specific}$) can be back-predicted from the crystal structure of the complex[81]. To obtain the value for the population of the encounter state (p), its ensemble-averaged PRE ($\Gamma_2^{encounter}$, eq 5.2), needs to be determined. This PRE can be calculated from the spatial distribution of the constituent protein-protein orientations. Recently, we have shown that MC simulations provide an accurate description of the wt Cc - CcP encounter complex, which is in excellent agreement with experimental data and thus can be used to derive the required $\Gamma_2^{encounter}$ (chapter 4)[184]. In the MC simulations, the proteins are treated as rigid bodies with partial charges assigned at the atomic level. Intermolecular electrostatic interactions and an exclusion grid, designed to prevent steric clashes, guide the

subsequent protein - protein docking. Using this approach, here we simulated ensembles of encounter orientations for the three mutant complexes.

As can be seen from the energy versus the distance root mean square (DRMS) (chapter 4)[23] plots, a single cluster of solutions is produced in all MC runs (Figure 5.9). The energy distributions of the MC solutions for R13K and T12A Cc - CcP are similar to that of the wt complex. For R13A Cc - CcP the distribution is shifted to higher energy values, indicating that removal of a positively-charged sidechain impairs protein association, as would be expected for an electrostatically-driven process.

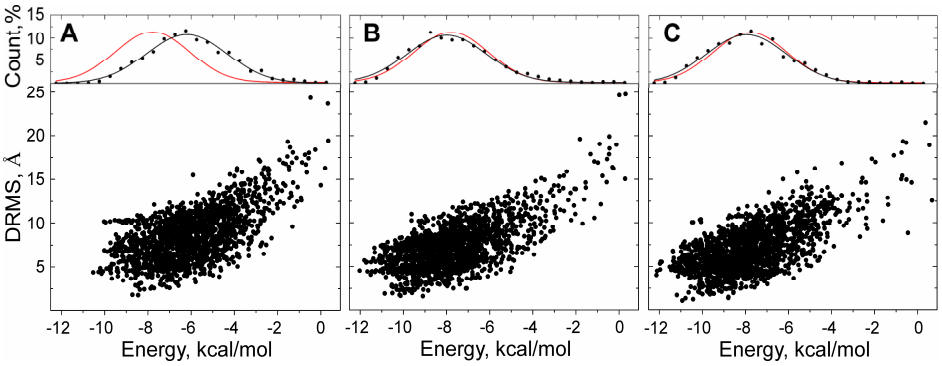


Figure 5.9. DRMS versus energy plots for the Monte Carlo simulations of the encounter state. (A) R13A, (B) R13K, and (C) T12A Cc - CcP. In each panel the upper graph shows the energy distribution fitted with a Gaussian with parameters (A) $r^2 = 0.98$ and $x_c = -6.2 \pm 0.1$ kcal/mol; (B) $r^2 = 0.98$, $x_c = -8.0 \pm 0.1$ kcal/mol); and (C) $r^2 = 0.97$, $x_c = -8.0 \pm 0.1$ kcal/mol. A Gaussian distribution for the wt complex (red; $r^2 = 0.98$, $x_c = -7.8 \pm 0.1$ kcal/mol) is given for comparison.

The spatial distributions of the simulated MC ensembles are shown in Figures 5.10-5.13 for the R13A, wt, R13K and T12A Cc – CcP complexes. The analysis and graphs were made by Dr. A Volkov. Similar to what was seen for the wt complex (chapter 4)[184], most MC solutions are found surrounding the dominant Cc - CcP orientation observed by X-ray crystallography[81] and solution NMR spectroscopy[19]. In all complexes, Cc explores a narrow patch on the CcP surface, navigating between two local energy minima (panels A and C in Figures 5.10-5.13), which were also observed in the classical Brownian dynamics study[102]. This patch contains the binding site of the specific complex, indicating that the electrostatic pre-orientation of CcP - occurring

during a diffusional approach of Cc - steers Cc to the dominant binding site. As it probes the CcP surface, Cc experiences considerable rotational motion, as evidenced by broad distributions of CcP centers of mass around Cc (panels D-F in Figures 5.10-5.13). Overall, the extent of the conformational space sampled by the proteins in the wt and variant encounter complexes appears to be very similar, implying that the introduced Cc mutations do not significantly alter the geometry of the encounter state.

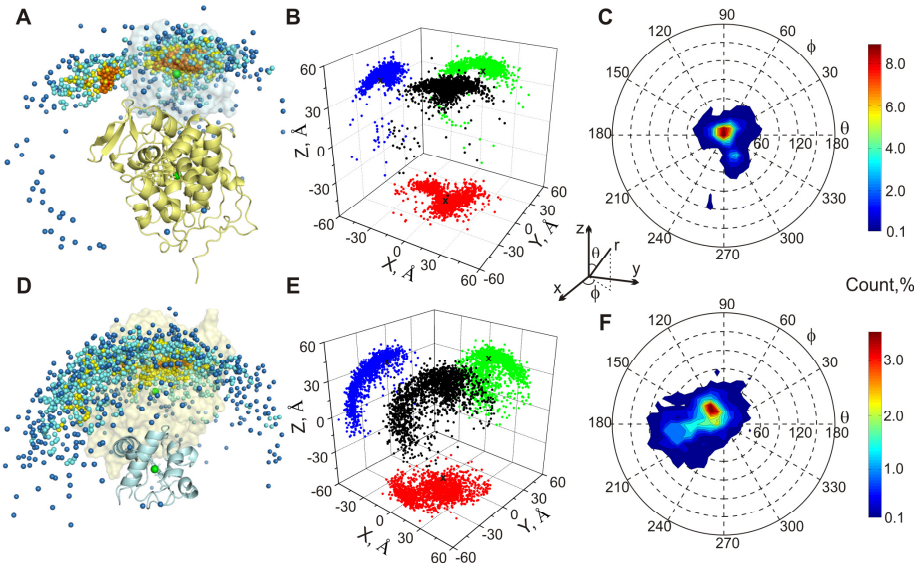


Figure 5.10. Monte Carlo simulation of R13A Cc - CcP encounter state. (A, D) Spatial distribution of the encounter complexes. Cc and CcP molecules in the specific complex are coloured light yellow and blue, respectively, with one protein in cartoon and the other as a transparent surface. The centres of mass of (A) Cc around CcP and (B) CcP around Cc in the simulated encounter ensemble are shown as spheres, coloured according to the density of the distribution increasing from blue to red. The highest densities denote the most favourable electrostatic orientations. Centres of mass of the proteins in the specific complex (green spheres) are aligned along the z-axis of the Cartesian coordinate system and correspond to spherical coordinates (θ, ϕ) = (0,0) as defined in the inset. The view in (D) is obtained by rotating that in (A) by 160° around y axis. (B, E) Distribution of (B) Cc centres of mass around CcP and (E) CcP centres of mass around Cc in the Cartesian coordinate space. Coloured symbols denote projections on the coordinate planes, and crosses indicate the position of the center of mass of the partner protein in the specific complex. (C, F) Distribution of (C) Cc centres of mass around CcP and (F) CcP centres of mass around Cc in the spherical coordinate system. The contour levels refer to the counts (in percent) of the MC output structures binned in 10° steps along both θ and ϕ dimensions.

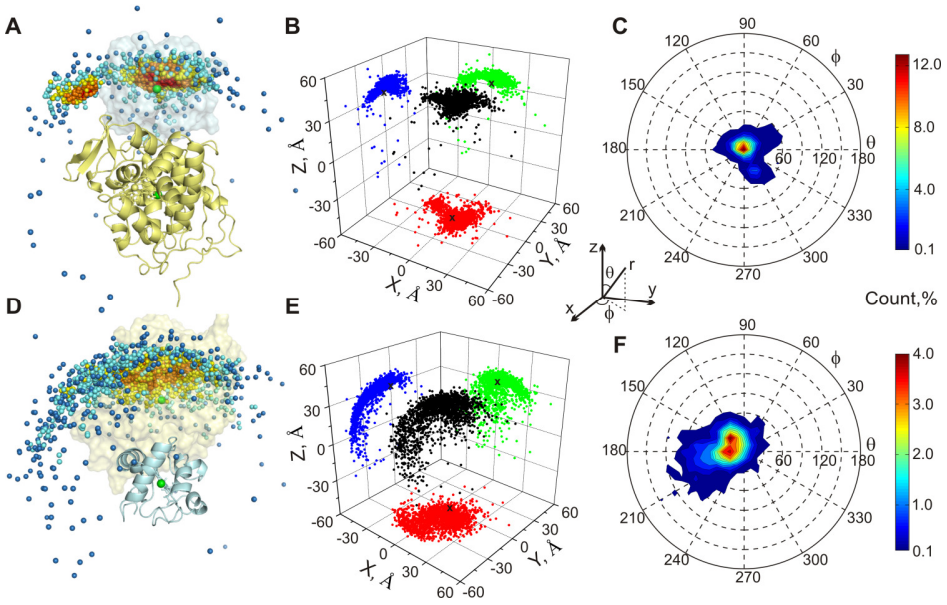


Figure 5.11. Monte Carlo simulation of wt Cc - CcP encounter state. (A, D) Spatial distribution of the encounter complexes and the distribution of the centers mass of (A-C) Cc around CcP and (D-F) CcP around Cc in (B, E) Cartesian and (C, F) spherical coordinate systems. See the legend to Figure 5.10 for details.

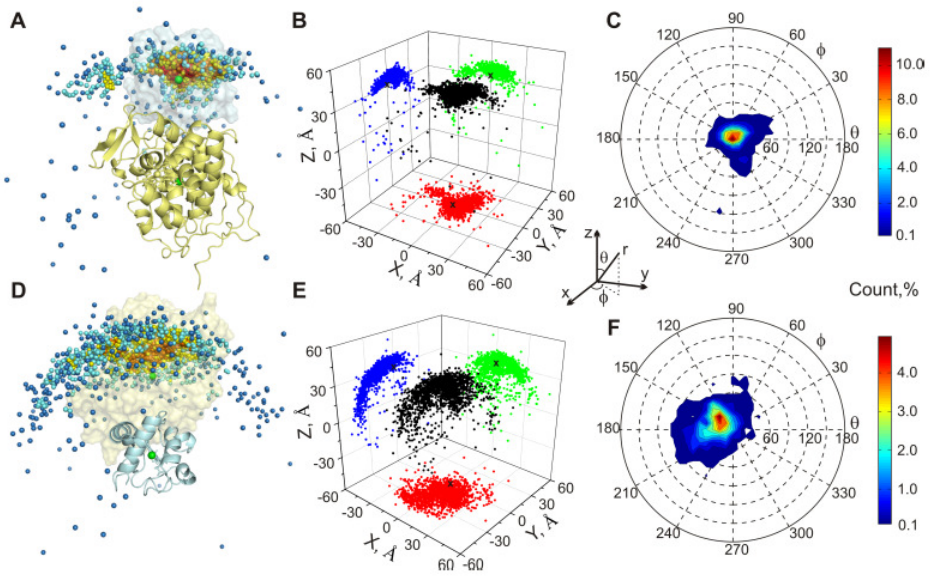


Figure 5.12. Mote Carlo simulation of R13K Cc - CcP encounter state. (A, D) Spatial distribution of the encounter complexes and the distribution of the centers mass of (A-C) Cc around CcP and (D-F) CcP around Cc in (B, E) Cartesian and (C, F) spherical coordinate systems. See the legend to Figure 5.10 for details.

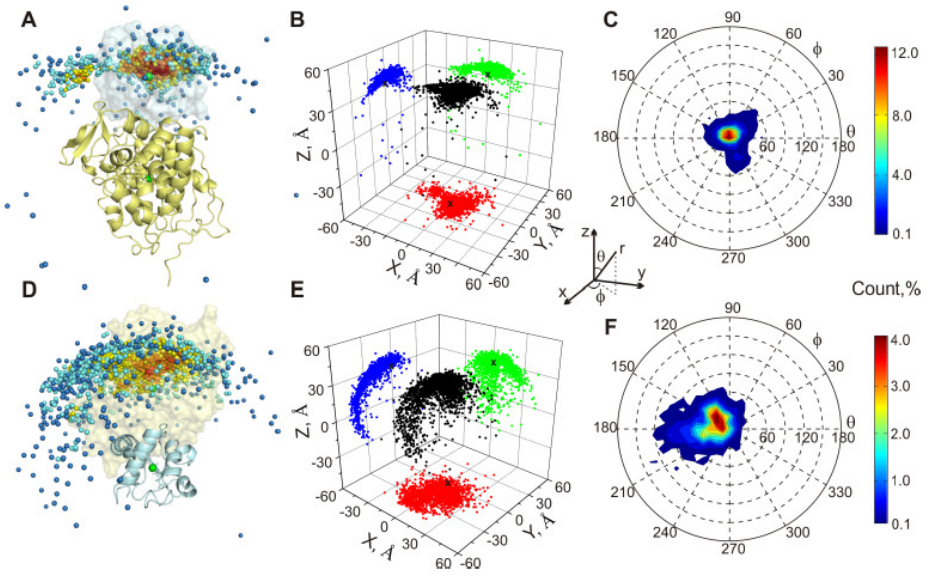


Figure 5.13. Monte Carlo simulation of T12A Cc - CcP encounter state. (A, D) Spatial distribution of the encounter complexes and the distribution of the centers mass of (A-C) Cc around CcP and (D-F) CcP around Cc in (B, E) Cartesian and (C, F) spherical coordinate systems. See the legend to Figure 5.10 for details.

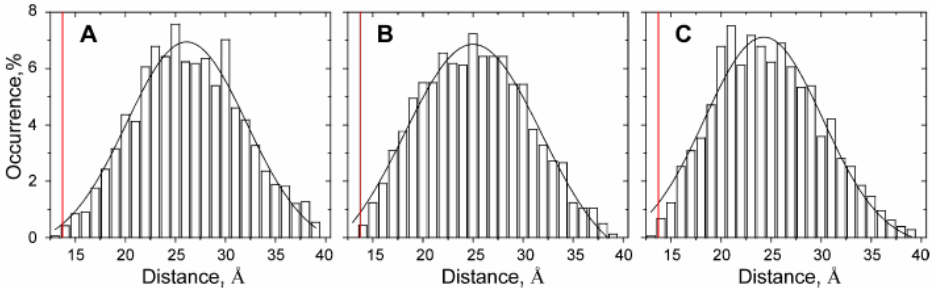


Figure 5.14. Intermolecular ET donor-acceptor distances in the encounter state. Edge-to-edge distances between the aromatic ring of CcP Trp 191 and haem group of Cc in the simulated encounter ensembles of (*A*) R13A, (*B*) R13K, and (*C*) T12A Cc - CcP. Black lines are Gaussian fits of the distributions with parameters (A) $r^2 = 0.95$ and $x_c = 26.1 \pm 0.2$ Å; (B) $r^2 = 0.97$, $x_c = 25.0 \pm 0.2$ Å; and (C) $r^2 = 0.95$, $x_c = 24.3 \pm 0.2$ Å. The red line marks the distance in the specific complex (13.7 Å).

Just as for wt Cc – CcP (chapter 4)[184], the intermolecular ET donor-acceptor distances in mutant encounters are large (Figure 5.14). At the observed donor-acceptor separations, the rate of intermolecular ET in the encounter state is expected to be very low[177]. This finding agrees well with the results of recent kinetics studies of Erman and co-workers, who concluded that most of the ET occurs in the dominant form of the Cc - CcP complex[181-183].

Encounter state populations

To obtain the encounter state population (p), the back-calculated PRE contributions of the specific and the encounter forms are fitted to the experimental Γ_2^{obs} (eq 5.2). The quality of the fit is estimated by calculating a Q factor (eq 5.3), and the p values at the lowest Q are selected (chapter 4)[184]. A single, well-defined minimum in Q = f(p) plots, corresponding to the optimal p values, is found for all studied complexes (Figure 5.15), and a combination of the specific form (p = 0) and the encounter state (p = 1) provides a better agreement with the observed PREs than either of these forms alone (Figures 5.16-5.18).

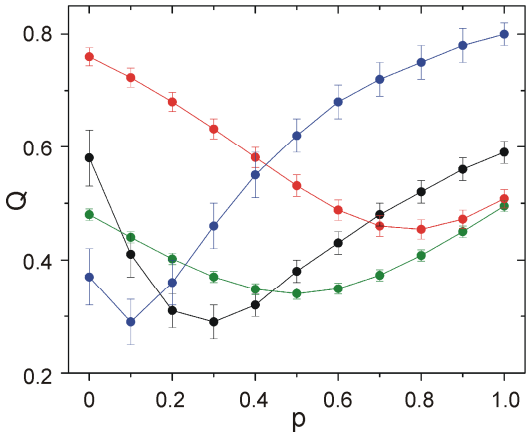


Figure 5.15. Q-factor analysis of the encounter complexes. Q factors for the PREs calculated at varying p for wt (chapter 4)[184] (black), T12A (blue), R13K (green), and R13A (red) Cc - CcP. The errors are standard deviations (chapter 4)[184].

Remarkably, there is a clear shift from p = 0.3 in the wt complex to p = 0.1 in T12A Cc - CcP and, in the other direction, to p = 0.8 in R13A Cc - CcP (Figure 5.15). In the latter complex, the encounter state is no longer a minor species but rather the dominant form. Thus, it appears that a single mutation can both decrease (T12A) and significantly increase (R13A) the population of the encounter state.

The crystallographic wt Cc - CcP orientation[81], known to be the dominant species in solution[19], was used to represent the specific form in all complexes. This approach is justified as 1) at the resolution of protein backbones, T12A mutation does not perturb the structure of the specific form of the complex[185]; and 2) although no structures of the weak R13K and R13A Cc - CcP complexes could be obtained, given the overall similarity of their PRE profiles (Figure 5.5A,B) to those of the wt[184], their specific binding geometries must not be greatly altered. Nevertheless, slight differences in the structures of the specific forms would give rise to higher Q factors observed for the R13-substituted complexes (Figure 5.15). Also, at higher encounter state populations, the quality of the agreement with the experimental data becomes sensitive to the level of detail for the simulated ensemble, which is limited by the approximations inherent in the simulation protocol.

Shifting the equilibrium between the encounter state and the specific form

There is a qualitative agreement between p and the size of the NMR chemical shift perturbations ($\Delta\delta_{avg}$) in Cc - CcP complexes (Figures 5.4 and 5.15). This finding strengthens the conclusions of earlier studies that correlated decreased $\Delta\delta_{avg}$ with an increased dynamics in transient ET protein complexes[26,29,124] and suggests that most of the binding shifts arise from the specific form of a complex. Interestingly, there is a linear correlation between p and the free binding energy (ΔG_B), with weaker complexes exhibiting higher p values (Figure 5.19). The T12A mutation increases the binding constant by an order of magnitude and decreases p to 0.1; R13K weakens the complex 4-fold and gives rise to equal populations of the specific and encounter forms; finally, R13A decreases the binding 100-fold and increases p to 0.8 (Table 5.1 and Figure 5.19).

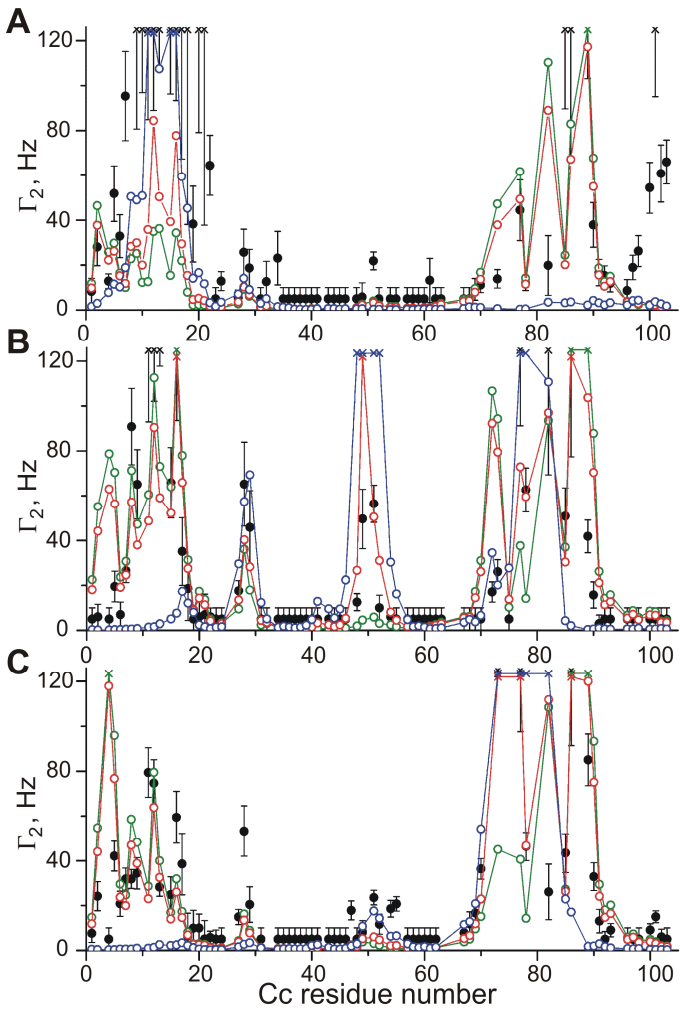


Figure 5.16. Observed and calculated PREs for R13A Cc - CcP-SL complexes. PRE profiles for SLs attached at position (A) N38C, (B) N200C and (C) T288C. Experimental Γ_2^{obs} (black) and PREs back-calculated for the specific orientation (blue), simulated encounter state (green), and the combination of the two (red) at the optimal p value of 0.8. Crosses indicate the values of $\Gamma_2 \geq 125$ Hz for the calculated PREs or identify the residues whose resonances disappear in the paramagnetic spectrum. For the latter only the lower limit of Γ_2^{obs} could be estimated from the spectral noise level. For residues with $I_{ox}/I_{red} \geq 0.9$ (Figure 5A), Γ_2^{obs} is set to 5 ± 5 Hz; otherwise, the errors are standard deviations.

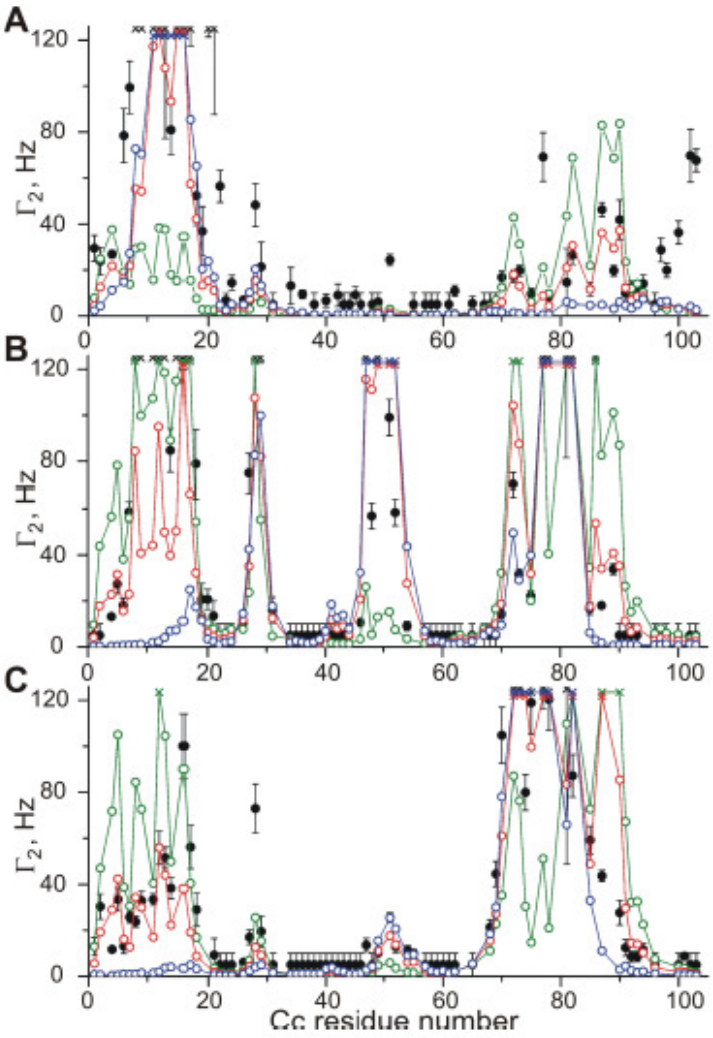


Figure 5.17. Observed and calculated PREs for R13K Cc - CcP-SL complexes. PRE profiles for SLs attached at position (*A*) N38C, (*B*) N200C and (*C*) T288C. Experimental Γ_2^{obs} (black) and PREs back-calculated for the specific orientation (blue), simulated encounter state (green), and the combination of the two (red) at the optimal p value of 0.4. See the legend to Figure 5.16 for more details.

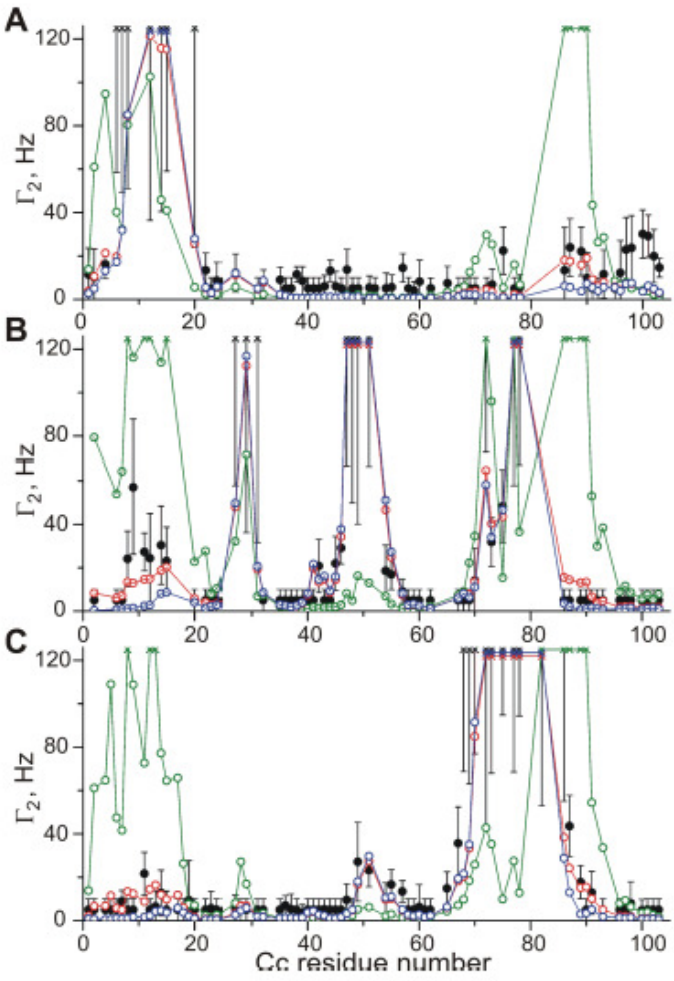


Figure 5.18. Observed and calculated PREs for T12A Cc - CcP-SL complexes. PRE profiles for SLs attached at position (*A*) N38C, (*B*) N200C and (*C*) T288C. Experimental Γ_2^{obs} (black) and PREs back-calculated for the specific orientation (blue), simulated encounter state (green), and the combination of the two (red) at the optimal p value of 0.1. See the legend to Figure 5.16 for more details.

When ΔG_B is decomposed into individual contributions of the two forms (see theory section), another trend can be discerned. The mutations exert a larger effect on the specific orientation than on the encounter state and, consequently, as complexes become progressively weaker, the contribution of the latter to ΔG_B steadily increases (Figure

5.19). In the extreme case of R13A, the encounter state becomes the dominant species and imparts most of the thermodynamic stabilization to the overall protein binding. Thus, the introduced mutations reshape the energy landscape of Cc - CcP interaction and, in case of R13A, can even interchange the energy levels of the specific and the encounter forms (Figure 5.19).

Earlier kinetic studies, mainly concerned with Cc F82 variants, already demonstrated that small changes in the interface can reduce the rate of intermolecular ET and cause Cc to assume different orientations in crystal structures[186,187]. These results show how easily the specific complex is perturbed, implying only a minor stabilization relative to the encounter state. A similar finding was reported in a study of the complex between cytochrome *f* and plastocyanin from *Prochlorothrix hollandica*: NMR data showed that the native complex exhibited dynamics, which was enhanced by flattening the plastocyanin interaction surface by mutagenesis[32]. Recently, Xiong et al.[31] converted a highly dynamic complex of myoglobin and cytochrome b_5 into a more static, specific one by introducing three charge-reversal mutations around the front face of myoglobin. In that case, reduction of the electrostatic repulsion between the proteins appears to be sufficient to form a specific complex.

Many experimental studies of protein-protein association (reviewed recently in ref.[9]) focused on the transition state (TS) of the free-energy barrier separating the encounter state and the specific form (AB* and AB in the schematic diagrams of Fig. 5.19) and believed to resemble the latter[9]. For example, it was shown that mutations that enhance the association rate in the TEM1-BLIP complex can make the TS less diffusive[153]. There is an essential difference between the TS and the encounter complex. By definition, the former is extremely short-lived and, due to its very low population, can only be probed by kinetic experiments. The encounter state – an intermediate state between free proteins and the specific form – can be populated significantly, which allows its characterization by PRE NMR spectroscopy. Here, we used this tool to address the entire ensemble of protein-protein orientations constituting the encounter state, thus providing an insight into the equilibrium between this state and the specific form of the complex. Our present results suggest that, in addition to its accepted role in enhancing the binding kinetics[9,17], the encounter can make a large contribution to the thermodynamic stabilization of weak protein interactions. This finding could rationalize

the existence of highly dynamic macromolecular complexes consisting entirely of multiple encounter forms[26,30].

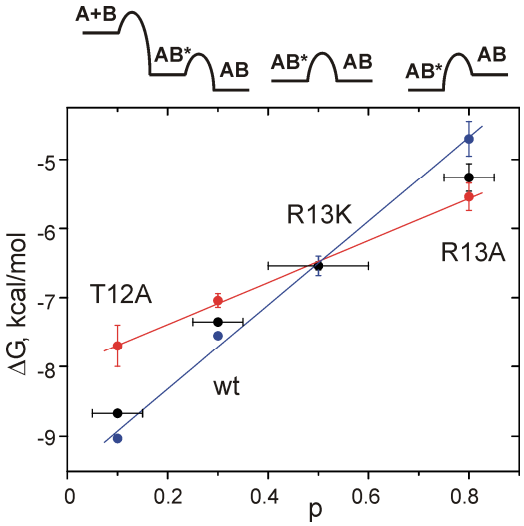


Figure 5.19. Energy contributions of the encounter states in variant Cc - CcP complexes. Experimental ΔG_B (black) decomposed into the contributions of the specific form (blue) and an encounter state (red). The former and the latter fit to linear trends with r^2 = 0.98, slope 6.1 kcal/mol and r^2 = 0.96, slope 3.1 kcal/mol, respectively. The errors are standard deviations derived from the uncertainties in experimental binding constants and determined p values (eq 5.14). The diagrams above the plot schematically outline relative positions of the energy levels of the free proteins (A+B), the specific form (AB) and an encounter state (AB*) for the Cc - CcP complexes.

On the specificity of protein-protein interactions

In the context of a single protein complex, specificity can be defined as an ability to adopt a unique, well-defined orientation as opposed to multiple, possibly overlapping, binding modes. We propose that p is a good metric for specificity in weak protein complexes, which can be used to classify a range of transient interactions varying from mostly single-orientation ($p < 0.5$)[19,20], to less specific ones ($p \sim 0.5$)[29,124], to those dominated by an encounter state ($p > 0.5$)[26,30]. In fact, as we show here, p can be modulated within a single protein complex, thereby shifting the delicate balance of the specificity in either direction.