

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/36430> holds various files of this Leiden University dissertation.

Author: Schilder, Jesika T.

Title: Making the invisible visible : paramagnetic NMR and the transient protein complex

Issue Date: 2015-11-24

MAKING THE INVISIBLE VISIBLE

PARAMAGNETIC NMR AND THE TRANSIENT PROTEIN COMPLEX

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Prof. Mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 24 november 2015
klokke 10:00 uur

door

Jesika T. Schilder

Geboren te Whitehorse, Yukon, Canada in 1986

Making the invisible visible - paramagnetic NMR and the transient protein complex

Jesika T. Schilder

Doctoral thesis, Leiden University, 2015

Cover image: Complex between yeast cytochrome *c* and yeast cytochrome *c* peroxidase (PDB 2GB8).⁽¹⁾

ISBN number: 978-94-6182-615-2

© 2015, Jesika T. Schilder

Printed by Off-Page – Amsterdam, Netherlands

Promotor: Prof. Dr. M. Ubbink
Promotiecommissie: Prof. Dr. G. Karlsson (University of Gothenburg)
Prof. Dr. S. de Vries[†] (Delft University of Technology)
Prof. Dr. J. Brouwer
Prof. Dr. M.H.M Noteborn
Dr. M. Huber

The research described in this thesis was performed in the Protein Chemistry Department of the Leiden Institute of Chemistry at the Universiteit Leiden in Leiden, the Netherlands.

The research described in this thesis was financially supported by the Netherlands Organization for Scientific Research (NWO), VICI grant 700.58.441.

“All that glitters may not be gold, but at least it contains free electrons.”

- *John Desmond Bernal*

For those with wild dreams and iron will....

TABLE OF CONTENTS

	List of abbreviations	1
	List of symbols	3
Chapter 1	Introduction to transient protein complexes	5
Chapter 2	Introduction to paramagnetic NMR and the cytochrome <i>c</i> – cytochrome <i>c</i> peroxidase complex & thesis outline	19
	Thesis outline	35
Chapter 3	NMR assignment and analysis of haem iron $\Delta\chi$ -tensors for two forms of yeast cytochrome <i>c</i> peroxidase: serendipitous discovery of time-dependent hydrogen fluoride binding	37
Chapter 4	The cytochrome <i>c</i> peroxidase and cytochrome <i>c</i> encounter complex: The other side of the story	65
Chapter 5	Protein docking using an ensemble of spin labels optimized by intra-molecular paramagnetic relaxation enhancement	85
Chapter 6	Ultra-weak self-association between cytochrome <i>c</i> peroxidase molecules observed by paramagnetic NMR	117
Chapter 7	Combining multiple types of paramagnetic NMR data to study the cytochrome <i>c</i> – cytochrome <i>c</i> peroxidase encounter complex	133
Chapter 8	General discussion, conclusions & perspectives	159
	Reference list	165
	Appendices	185
	English summary	193
	Nederlandse samenvatting	197
	List of publications	202
	Curriculum vitae	203

LIST OF ABBREVIATIONS

AV	average violations
BD	Brownian dynamics
BEST	band-selective excitation short-transient
BMRB	biological magnetic resonance bank
CAM	chloramphenicol
Cc	cytochrome <i>c</i>
CcP	cytochrome <i>c</i> peroxidase
CLaNP	caged lanthanoid NMR probe
CpdI(II)	CcP compound I (II)
CSA	chemical shift anisotropy
CSP	chemical shift perturbation
cw	continuous wave
CYC1	gene for yeast <i>iso-1</i> Cc
CYC3	gene for Cc haem lyase
cyt	cytochrome
DD	dynamic docking
DRMS	distance based root mean square deviation
EDTA	ethylenediaminetetraacetic acid
EIN	N-terminal domain of Enzyme I
EPR	electron paramagnetic resonance
ET	electron transfer
HPr	histidine-containing phosphocarrier protein
HS	high-spin
HSQC	heteronuclear single quantum coherence
KAN	kanamycin
LS	low-spin
MC	monte carlo
MTS	1-acetoxy-2,2,5,5-tetramethyl- δ^3 -pyrroline-3-methyl)-methanethiosulfonate
MTSL	1-oxy-2,2,5,5-tetramethyl-2,5-dihydropyrrol-3-ylmethyl methanethiosulfonate
NaPi	sodium phosphate
NMR	nuclear magnetic resonance spectroscopy
NOE	nuclear Overhauser effect
Pc	plastocyanin
PCS	pseudocontact shift
PDB	protein databank

PLN	phospho-lamban
PMSF	phenylmethanesulfonylfluoride
pNMR	paramagnetic nuclear magnetic resonance spectroscopy
PPIX	iron-free protoporphyrin IX haem precursor
PRE	paramagnetic relaxation enhancement
pyMTSL	3-methanesulfonilthiomethyl-4-(pyridin-3-yl)-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-1-yloxy
RACS	residual anisotropic chemical shifts
RDC	residual dipolar coupling
RMSD	root mean square deviation
RS	resting state
SD	simple docking
SL	spin label
TROSY	transverse relaxation optimized spectroscopy
vdW	Van der Waals
WT	wild type
$\Delta\chi$ -tensor	magnetic susceptibility tensor

LIST OF SYMBOLS

d_{ij}^{ens}	distance matrix from mean ensemble structure
d_{ij}^{xray}	distance matrix from mean X-ray crystal structure
A_{280nm}	absorbance at 280 nm
A_{408nm}	absorbance at 408 nm
B_0	magnetic field
D	zero-field splitting constant
dis^{calc}	back-calculated distances
dis^{obs}	Observed distances
D^{res}	calculated RDC
f_{bound}	fraction bound
g_e	electronic g-factor
h	Plank's constant
I_{dia}	peak intensity in the diamagnetic spectrum
I_{para}	peak intensity in the paramagnetic spectrum
I_{para}/I_{dia}	intensity ratio
k	Boltzmann constant
K_B	binding constant
K_D	dissociation constant
k_{ex}	chemical exchange rate
Ln^{3+}	lanthanoid ion
N	total number of i, j pairs
N_{e2}	second nitrogen of histidine side chain
$R_{1,Curie-spin}$	longitudinal relaxation rate due to Curie mechanism
$R_{1,dipole}$	longitudinal relaxation rate due to Solomon mechanism
$R_{1,para}$	paramagnetic contribution to the longitudinal relaxation rate
$R_{2,Curie-spin}$	transverse relaxation rate due to Curie mechanism
$R_{2,dia}$	diamagnetic contribution to the transverse relaxation rate
$R_{2,dipole}$	transverse relaxation rate due to Solomon mechanism
$R_{2,para}$	paramagnetic contribution to the transverse relaxation rate
r_{AB}	distance between J-coupled nuclei A and B
r_{IM}	distance between nucleus I and paramagnetic centre
S	spin quantum number
S_8	methionine sulphur
T	absolute temperature

T_0	initial time point
T_1	Second time point
β	Bohr magneton
γ_A	gyromagnetic ratio of nucleus A
γ_B	gyromagnetic ratio of nucleus B
γ_I	gyromagnetic ratio of nucleus I
$\Delta\chi_{ax}$	axial component of χ -tensor
$\Delta\chi_{rh}$	rhombic component of χ -tensor
δ_{PCS}	pseudocontact shift difference
$\Delta\delta_{avg}$	average CSP
$\Delta\omega$	difference in resonance frequency
μ_0	vacuum permeability
τ_c	$\tau_c^{-1} = \tau_r^{-1} + \tau_s^{-1}$
τ_r	rotational correlation time
τ_s	longitudinal electronic relaxation time
ω_I	Larmor frequency of nucleus I

CHAPTER 1

Introduction to transient protein complexes

Based on the review article: **Schilder J.** and Ubbink M. (2013). *Formation of transient protein complexes*. *Curr Opin Struct Biol* **23**(6): 911-918.

ABSTRACT

The encounter complex of two proteins is a dynamic intermediate state that guides proteins to their binding site, thus enhancing the rate of complex formation. They are particularly useful for complexes that must balance a biological requirement for high turnover with the need for specific binding, such as electron transfer (ET) complexes. Here, we describe the current methods for studying and visualizing encounter complexes. We discuss recent developments in mapping the energy landscapes, the role of hydrophobic interactions during encounter complex formation and the discovery of futile encounter complexes. These studies have not only provided insight into encounter complexes of ET proteins, but also opened up new questions and approaches for studying encounter complexes in other weakly associated proteins

INTRODUCTION

Historically, protein complex formation has been viewed as a simple two state process in which the proteins are either free and unbound or bound in a static, specific complex. However, this view has changed to include an intermediate stage, commonly known as the encounter complex (Figure 1.1A-C).

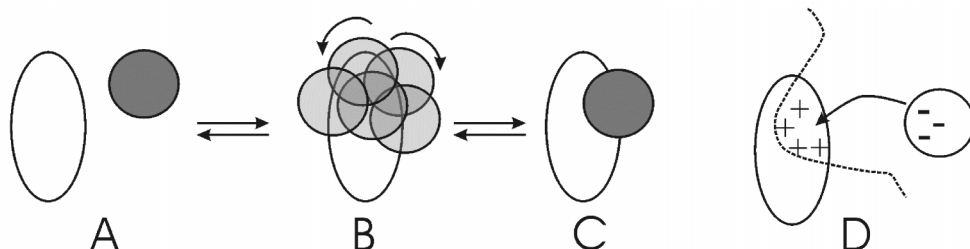


FIGURE 1.1 Model of protein complex formation. **A-C**) During complex formation, free proteins (A) first form a weakly bound transient intermediate, the encounter complex (B), before proceeding to the final specific complex (C). **D**) As the binding partners approach each other, long-range electrostatic interactions between charged residues on the protein surfaces pre-orient the proteins for complex formation. Reprinted with permission from ⁽²⁾. © 2012 Portland Press Limited.

The encounter complex consists of an ensemble of low energy, weakly associated conformations that are in a dynamic equilibrium with a specific complex. In this loosely associated state, the proteins are free to rotate and reorient themselves, sampling each other's surfaces and increasing the number of contacts, until optimal binding geometry is reached and the complex can proceed to the tightly bound, active state.⁽³⁾ The encounter complex is stabilized mainly by long-range electrostatic interactions and desolvation forces, with short-range interactions generally becoming important later in specific complex formation.⁽⁴⁾ The electrostatic forces start to work already when the proteins approach each other, resulting in a pre-orientation in the electric field (Figure 1.1D). If the charges are placed correctly on the surface, this process increases the chance that the proteins collide with the binding sites in close approximation. The charge attraction also prolongs the lifetime of the encounter. Thus, the encounter complex can aid in specific complex formation by reducing the surface area on the two proteins that needs to be searched before the binding site is found as well as by extending the lifetime of diffusional collisions between the two proteins.⁽⁵⁾ However, a consequence of charged patches on a protein surface is that the partner can bind in many

different orientations with more or less the same decrease in free energy^{1,6)} For formation of a specific complex, a single orientation is required with a much lower free energy than that of similar orientations, which is usually achieved by multiple short-range interactions (van der Waals, hydrogen-bonding, hydrophobic contacts and specific salt bridges). Consequently, stabilisation of the encounter complex enhances complex formation, but at the same time can counteract the formation of a specific complex. This dilemma is particularly relevant for weak complexes, such as formed between ET proteins.

The encounter state plays an important role in complexes in which the biological requirement of a high turnover rate must be counterbalanced with the demand of forming a specific interaction, which is the case for many ET proteins. This review will mainly discuss ET complexes, because much work has been done to study their encounter states, although examples of other, higher affinity complexes will also be given. Interestingly, the nature of the encounter complex varies for each ET complex, depending on the exact physiological requirements of the protein complex. In particular, the fraction of the proteins bound in the encounter state versus the productive state can differ markedly between complexes, depending on the specificity of binding in the productive state.⁷⁾ For example, the encounter complex for cytochrome *c* (Cc)-cytochrome *c* peroxidase (CcP) complex comprises 30% of the population^{1, 8)} and this population can readily be shifted with targeted point mutations at the interface of the specific complex to as low as 10% or as high as 80%.⁹⁾ Similar results have been shown for the myoglobin-cytochrome (cyt) *b*₅ complex¹⁰⁻¹²⁾ and the plastocyanin (Pc)-cyt *f* complex.¹³⁾ Furthermore, some complexes have been shown to exist purely in a productive encounter complex, never proceeding to a specific complex, such as the adrenodoxin-Cc complex¹⁴⁾ and myoglobin-cyt *b*₅ complex.¹⁵⁻¹⁸⁾

This paper will give a brief overview of the current methods for studying and visualizing ET encounter complexes. It will also mention the latest developments in mapping the energy landscapes of these associations as well as discuss the increasing recognition of the role that hydrophobic interactions play in encounter complex formation. Finally, complexes that are not optimized for fast complex formation, resulting in ‘futile’ encounters, will be considered.

¹This type of binding was dubbed the Velcro model of binding,⁵⁾ because it resembles pieces of Velcro, which can bind with similar strength in many orientations. However, that metaphor is also deceptive, because Velcro halves need to be separated completely before another orientation can be formed, whereas it is believed that the proteins in the encounter complex constantly rearrange within the complex without complete dissociation, although no direct evidence for that assumption is available.

METHODS TO STUDY ENCOUNTER COMPLEXES

KINETICS

During kinetic experiments, some form of spectroscopic change is measured in a time-resolved manner. This change can be a consequence of a chemical reaction, such as an ET reaction. Under certain conditions the observed rate depends not (only) on the ET reaction itself but is influenced by the type of complex formed or the rate of complex formation; thus, indirectly, information about the encounter complex can be extracted. In this manner, stopped flow experiments have been used to study the effects of changes in protein surface charges as well as buffer ionic strength on the ET rate in the cyt *f*-Pc complex. This complex was shown to exist in multiple conformations⁽¹⁹⁻²²⁾ including minor states that had not been observed in previous nuclear magnetic resonance spectroscopy (NMR) studies.⁽²³⁾ Likewise, transient absorption kinetics measurements have been used to study the complex of zinc-substituted myoglobin and cyt *b*₅. Using flash photolysis to observe triplet-quenching and ET, the complex was shown to exist purely in the encounter state.^(18, 24)

The association rate of complex formation can also be detected directly, if it leads to a change in tryptophan fluorescence. In combination with mutagenesis studies it has been possible to characterize interactions between specific residues in the encounter complexes of several proteins.⁽²⁵⁾ For kinetic data to be interpreted with respect to the encounter complex in such cases, the complex formation must be considered as an intramolecular chemical reaction. So strictly speaking, it is the transition state (peak in the energy landscape) between the encounter and specific states that is probed in such kinetic experiments, whereas in the structural studies discussed below, the wells in the energy landscape are studied. Because the transition state is closer to the specific complex than the encounter state is, kinetic and structural studies are complementary.

NMR STUDIES

Various NMR observables can be used to characterize encounter states. However, it should be realized that each of these represents an average of all the states of the protein complex in equilibrium, because exchange between these different conformations usually occurs much faster than the NMR time scale.⁽⁵⁾ Nevertheless, it has been shown that the magnitude of backbone amide chemical shift changes upon complex formation is a good indicator for the degree of dynamics within the protein complex, and thus the population of

the encounter complex, at least if it is dominated by electrostatic interactions. A possible explanation for this observation is that in the encounter complex the proteins remain solvated and thus the chemical environment of the amides hardly changes, resulting in very small perturbations, while in the specific complex, the solvent layer is changed more drastically, causing large perturbations of the amide resonances.^(15, 26, 27)

More recently, paramagnetic NMR (pNMR) has proven to be a powerful technique for studying lowly populated states such as the encounter complex. Paramagnetic effects are generated by unpaired electrons in paramagnetic centres that change the resonance frequency or the relaxation rate of nearby nuclei.⁽⁷⁾ Paramagnetic relaxation enhancement (PRE) falls off with the sixth power of the distance between the paramagnetic centre and observed nucleus, and can affect nuclei within 10-40 Å of the paramagnetic centre. This extreme distance dependence makes PRE NMR particularly sensitive to minor populations, allowing areas on the protein surface that are visited by the partner protein in the encounter complex to be delineated.^(1, 28, 29) Also pseudocontact shifts (PCS)⁽¹³⁾ and paramagnetically induced residual dipolar couplings (RDC)^(30, 31) are sometimes used to characterize encounter complexes, but these will be discussed further in Chapter 2.

VISUALIZING THE ENCOUNTER COMPLEX

Visualization of the encounter complex offers several challenges. Although kinetic data do not yield structural information directly, when combined with mutagenesis they can provide detailed information about where and how proteins interact. The observed interactions between specific residues can then be used as criteria to select complexes from protein docking simulations. However, due to the limited number of observed contacts, only an ensemble of structures potentially representing the encounter/transition states can be calculated.⁽²⁵⁾ NMR data provide more direct structural information, but they represent an average of all conformations present. Again, this presents an inverse problem in which many solutions can be found that fit the data. The end result is a cloud of low energy conformations, any combination of which could represent the encounter complex structure as the true number of conformations in the encounter complex ensemble is unknown. The maximum occurrence method can be used to identify which of these conformations are most likely to be sampled by the proteins.^(32, 33) However, the most conclusive result from PRE studies is an exclusion map. If a paramagnetic centre on one protein causes no PRE on the partner protein, then the partner does not visit that part of the protein surface, or at least so infrequently that no PRE is

generated. By using multiple centres at different sites on the protein, an exclusion map can be generated.^(1, 3, 8, 28, 34)

Several computer modelling methods have been used to visualize the encounter complex. First, explicit ensemble refinement involves simultaneously docking of several copies of one protein onto the other.^(28, 35) In this approach, many such dockings are performed yielding a cloud of conformations, combinations of which fit the experimental data. In theory, this should represent a negative image of the exclusion map, provided sufficient sampling of the protein surface with paramagnetic centres has been performed. A similar cloud can be obtained with a more simplistic approach which we call empirical ensemble simulations. Here, one starts with a previously determined structure of the complex. Then a series of random rotations and translations over the surface of the partner are performed, increasing the amplitude until an ensemble is generated that matches the data. This approach is useful when modelling the minimal degree of freedom within a protein complex required for agreement with observed averaging of PCS or RDC.^(13, 14) These two methods are essentially fitting procedures and do not include prior assumptions about the encounter complex.⁽⁷⁾ Brownian dynamics (BD) or Monte Carlo (MC) simulations are also commonly used.^(1, 3, 8, 10, 11, 25, 36) These methods simulate translational and rotational Brownian motion of the proteins as they are docked using only/mostly electrostatic forces. These simulations do not include internal conformational fluctuations and therefore can simulate much longer timescales than is currently possible with molecular dynamics simulations.⁽³⁷⁾ In such studies, the experimental data are used to guide the selection of generated models to obtain a description of the encounter complex (Figure 1.2).

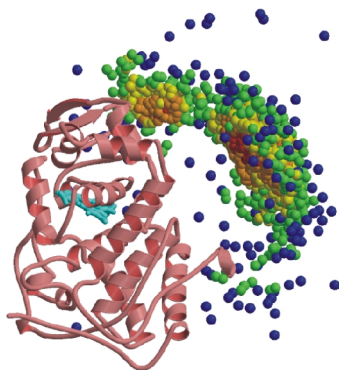


FIGURE 1.2 Simulation of the Cc-CcP encounter complex. CcP is represented by red ribbon with the haem group in cyan. The spheres represent the centre of mass of Cc after docking onto CcP using rigid-body MC simulations and are coloured by density, decreasing from red to blue. Reprinted with permission from ⁽⁸⁾. © 2010 American Chemical Society.

THE FREE ENERGY LANDSCAPE

The energy landscape of the encounter and final complexes is necessarily extensive, because the encounter complex consists of many orientations. In 1988, Northrup *et al.* published a seminal paper on the Cc-CcP encounter complex using theoretical BD simulations. The proteins were allowed to associate based on Coulombic electrostatic interactions for relatively long periods of time, while the proteins simultaneously reoriented themselves favourably for electron transfer. Several free energy minima were found in the resulting energy landscape, suggesting that several orientations of an encounter complex were formed rather than a single specific complex.⁽³⁸⁾

In 1999, Camacho *et al.* also used computer simulations to visualize the energy landscape for several protein pairs. The landscape was computed using semi-Coulombic electrostatic and empirical desolvation terms and the encounter complexes occurred at the interaction free energy minima. These theoretical approaches suggested that electrostatic steering plays a dominant role in encounter complex formation for highly charged proteins. It was further argued that once the binding partner has been steered into the broad binding pocket, short range electrostatics and desolvation could further direct the partner into the free energy minimum.⁽³⁹⁾ A later study suggested that for proteins with little or no net charge, several weakly specific pathways exist in the free energy landscape, driven mainly by desolvation forces, which guide the proteins into well-oriented encounter complexes.⁽⁴⁰⁾ This knowledge was utilized to develop a tuneable docking method that could account for the different driving forces at the different length scales along the binding pathway. First, electrostatic and desolvation forces dominate for complexes more than four angstrom away from the specific complex. This is followed at closer range by an increasing fraction of van der Waals (vdW) forces to find the final energy minima.⁽⁴¹⁾

In 2005, Spaar and Helms used BD simulation trajectories to make a theoretical model for the energy landscape of the classic barnase-barstar encounter complex. At several points along the binding trajectory, the positions and orientations of the proteins relative to each other were recorded and used to build occupancy maps, which were interpreted as a probability distribution for the calculation of the entropy landscape. The entropy landscape was combined with electrostatic and desolvation energies to generate the final energy landscape, with a clear path along the free energy minima.⁽⁴²⁾ This analysis shows which part in the energy landscape the encounter complex occupies and it can be used to study successful binding pathways or differences in association/dissociation pathways as well as the effects of mutations.⁽⁴³⁾

However, this view of the energy landscape as a simple funnel towards the bound state does not work well for small ET proteins. They can successfully transfer electrons in multiple orientations, provided the transfer distance is small enough, which allows their complexes to be highly dynamic.^(8, 44) This results in a relatively flat energy landscape, with no clear distinction between the specific and non-specific complexes, allowing rapid transition between all states.⁽⁴⁵⁾ An excellent example of such a complex is the myoglobin-cyt *b*₅ complex. In 2002, Liang *et al.* published two important papers in which they studied the interaction and ET rate between cyt *b*₅ and zinc substituted myoglobin. They found the complex to be highly dynamic and to assume multiple orientations. These orientations were very similar in energy but only a fraction were active in ET; this new type of binding was termed “dynamic docking” (DD) (Figure 1.3).⁽¹⁶⁾

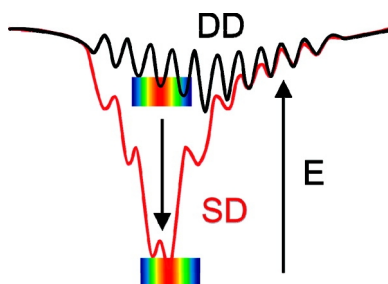


FIGURE 1.3 Comparison of energy landscapes generated using simple docking (SD) or dynamic docking (DD). The rainbows represent complexes capable of electron transfer. Reprinted with permission from ⁽¹¹⁾. © 2010 American Chemical Society.

Interestingly, changing the charge distribution on myoglobin by replacing the haem group with an analogue in which the negatively charged propionate groups had been neutralized can increase the ET rate. This result was interpreted as a change in the distribution of cyt *b*₅ orientations around myoglobin, leading to an increased fraction of ET active complexes.⁽¹⁷⁾ The same effect can be achieved with point mutations on the surface of myoglobin.^(10-12, 46, 47)

HYDROPHOBICITY AND THE ENCOUNTER COMPLEX

Despite the importance of hydrophobic interactions for encounter complexes, at least for weakly charged proteins as suggested by the modelling studies mentioned above, modelling encounter complexes using only electrostatic interactions successfully explained experimental results for the ET complexes of myoglobin and cyt *b*₅^(17, 46) as well as Cc and CcP.⁽⁸⁾ As of yet, the involvement of hydrophobic interactions in the encounter complex has only been reported

experimentally for the bacterial phosphotransferase system,⁽³⁾ during protein folding of an intrinsically disordered protein upon binding⁽⁴⁸⁾ and recently for the ET complex between cyt *f* and Pc.

The cyt *f*-Pc complex is involved in oxygenic photosynthesis in plants, algae and cyanobacteria. Pc accepts electrons from the haem group of cyt *f*, which is anchored to the thylakoid membrane via a single α -helix. The redox sites in both proteins are surrounded by a patch of hydrophobic surface residues, which form the binding site in the specific complex.⁽⁷⁾

Early NMR studies of the spinach Pc and turnip cyt *f* complex showed that the complex was quite specific and that both electrostatic and hydrophobic interactions occurred.^(49, 50) Similar studies on the same complex from the mesophilic cyanobacterium *Phormidium laminosum* showed that complex formation is much less dependent on ionic strength than the plant complex and that only the hydrophobic patches interact, suggesting that complex formation is controlled mainly by hydrophobic interactions.⁽²³⁾ Subsequent kinetic studies also demonstrated that electrostatics play only a minor role in the interaction.^(19, 20)

The Pc-cyt *f* complex from the cyanobacterium *Nostoc* sp. PCC 7119 was initially shown to be well-defined with hydrophobic interactions between the two hydrophobic patches as well as electrostatic interactions between the charged residues surrounding the patches on each protein.⁽⁵¹⁾ Kinetic studies of mutants showed that charge mutations on Pc had a much larger effect on the ET rate than similar mutations on cyt *f*, suggesting that Pc approaches with one face and then searches the cyt *f* surface for the binding site.^(52, 53) In a recent study, experimental PRE data were compared to back-calculated PRE data from the solution structure mentioned above but the data could not be fit. This signified that the complex was more dynamic than previously thought and a large proportion of the complex must exist in the encounter state.⁽⁵⁴⁾ A large follow-up study enabled visualisation of the encounter complex (Figure 1.4A) and demonstrated the involvement of both electrostatic and hydrophobic interactions. Note that such ‘clouds’ of orientations remain an approximation of the true encounter complex, because the number of restraints is insufficient to describe the entire ensemble (see above).

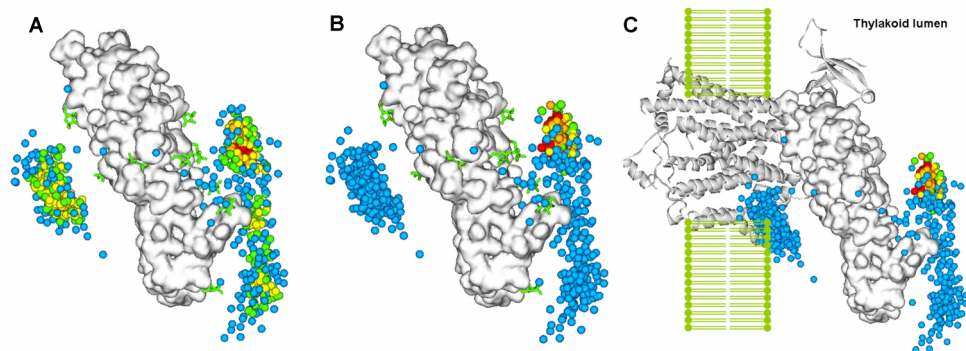


FIGURE 1.4 The encounter complex between Pc and cyt *f* from *Nostoc* sp. PCC7119. Cyt *f* is shown in white with the locations of the spin labels as green sticks. The spheres represent Pc centres-of-mass when docked onto cyt *f*. **A**) The spheres are coloured according to density, decreasing from red to blue. **B**) The spheres are coloured according to the distance between the Cu in Pc and the Fe in cyt *f*, increasing from red to blue (red ≤ 16 Å; orange ≤ 18 Å; yellow ≤ 20 Å; green ≤ 22 Å; blue > 22 Å). ET only occurs at a biologically relevant rate when this distance is < 16 Å. **C**) The Pc-cyt *f* complex is superimposed on the cyt *b₆/f* complex (PDB entry 2ZT9⁽⁵⁵⁾; shown as white ribbons), which is embedded in the thylakoid membrane. The spheres are coloured as in B. Reprinted with permission from ⁽⁴⁵⁾. © 2013 American Chemical Society.

The experimental data indicated that long-range electrostatic interactions pre-orient the unbound proteins for complex formation, enabling hydrophobic contacts to form. On the basis of this model, a smooth transition to ET active conformation(s) (i.e. without high-energy transition states) can then be envisaged via a gradual increase in the hydrophobic overlap.⁽⁴⁵⁾ If hydrophobic contacts are important for encounter complex formation, the distinction between the encounter complex and the specific complex blurs. This is in line with the view that multiple orientations can represent the active form in ET complexes because the dominant factor for ET is the distance between the redox centres. Any orientation of the partners in which the distance is short enough can lead to ET (Figure 1.4B).

FUTILE COMPLEXES

Not every encounter complex formed has the ability to proceed to an active, fruitful complex; thus, the idea of the futile encounter complex has recently emerged. In the description of the energy landscape of the barnase-barstar encounter complex, it was found that the energetically most favourable region was divided between two potential encounter complexes; one above the barnase binding interface, which leads to the active complex, and a second site near the RNA binding loop. This second site was not essential for complex formation but could help steer barstar towards the orientation required for fruitful encounter

complex formation. Key mutations were also shown to dramatically alter the free energy landscape and change the populations of fruitful versus futile encounter complexes.⁽⁴³⁾ Similar studies on the complex between TEM1- β -lactamase and β -lactamase-inhibitor proteins showed that the reaction rate is strongly influenced by the ability of a fruitful encounter complex to form, which in turn is strongly influenced by electrostatic optimization.⁽⁵⁰⁾

PNMR has provided similar results for the complex between the N-terminal domain of Enzyme I (EIN) and the histidine-containing phosphocarrier protein (HPr), which are involved in bacterial signalling. In 2010, Fawzi *et al.* used PRE data to map two distinct encounter complexes. The fruitful encounter complex was found near the EIN active site and was shown to be in equilibrium with the specific complex. A second HPr molecule was also bound elsewhere on the EIN surface, leading to a ternary complex. Although HPr bound at this site may be considered a futile encounter complex, binding at this site may also serve as an antechamber for HPr to allow quick successive, specific association once the phosphotransfer to the first HPr molecule is complete.⁽³⁴⁾

In the recent study of the *Nostoc* cyt *f* and Pc complex discussed above, futile encounter complexes have also been identified. The surface of cyt *f* sampled by Pc extends from the binding site at the haem all the way to the end of the small domain of cyt *f*. Many orientations of Pc are thus quite far from the site where ET can occur (Figure 1.4B) and it can be expected that the further away the Pc binds, the smaller the chance that it will proceed to form an ET active complex before it dissociates. The chance of being a futile complex thus increases with the distance from the ET active binding site. Therefore, the distinction between futile and fruitful may not be clear-cut but rather a continuum. Finally, a non-biologically relevant futile encounter complex was also been identified at the back of cyt *f* (Figure 1.4C). In the cell, cyt *f* is attached to the membrane via a single α -helix but during most experiments the protein is studied in a soluble form. Therefore, the futile complex identified at the back of cyt *f* cannot occur in the cell, as this site would be blocked by the membrane. This serves as a reminder that futile complexes need to be interpreted in a cellular context.⁽⁴⁵⁾

FUTURE PERSPECTIVES

Recent studies on the encounter complex have transformed the way we look at protein complex formation. The old two-state model has been replaced by a more complicated process involving a weakly-associated intermediate that can pre-orient the binding partners and enhance the rate of specific complex formation. Encounter complexes have been described for several ET complexes ranging from relatively well-defined, quickly proceeding to the active complex, to complexes that appear to exist solely in the encounter state. Furthermore, recent experimental evidence has provided support for the theoretical models that indicate that the formation of these encounter complexes can be driven not only by electrostatic interactions but also by hydrophobic interactions.

Encounter complexes often represent only a small fraction of the population in a given sample and consist of many orientations, making them difficult to study. This has led to the development of clever and highly sensitive kinetic and NMR techniques but the resulting data still present an inverse problem resulting in an ensemble of possible solutions instead of a precise description of the entire complex. Nonetheless, current methods have allowed us to model several associations between different proteins as well as reshaped our understanding of protein interaction energy landscapes.

Encounter complexes have been shown to play a key role in complex formation for ET proteins. Although several non-redox encounter complexes have been studied in depth, a broader survey of the role that encounter complexes play in other transient complexes, such as cell signalling proteins, would be of great interest. Furthermore, few studies have looked at the role of encounter complexes in intrinsically disordered proteins. Finally, the recent ability to observe futile encounter complexes has opened up a new set of questions and further research will be required to determine how common they are and if they truly are futile.

CHAPTER 2

Introduction to paramagnetic NMR spectroscopy
and the cytochrome *c* – cytochrome *c* peroxidase
complex

ABSTRACT

Paramagnetic NMR spectroscopy utilizes the effects generated by unpaired electrons of paramagnetic centres to study protein structure and dynamics. It relies on three main effects: PRE, which is highly distance dependent and is therefore incredibly sensitive to lowly populated states; PCS, which can be observed over large distances (≥ 60 Å) and provide both distance and orientation information; and RDC, which provide orientation information independent of the distance between the paramagnetic centre and the observed nuclei. These effects can be converted into restraints that can be used independently or in conjunction for structural modelling. They are particularly well-suited for studying highly dynamic, transient protein complexes, such as ET complexes, whose lowly populated states are crucial for a variety of biological processes but are practically invisible to conventional structural biology techniques. This chapter will briefly describe the theory behind pNMR as well as provide background information on the protein complex used in this study, the Cc-CcP complex.

INTRODUCTION

As described in chapter one, protein complex formation cannot simply be considered a two-state process in which the partners proceed directly from an unbound to bound state. Often the process proceeds via a weakly bound intermediate called the encounter complex that consists of an ensemble of highly dynamic, low energy conformations.^(3, 57) While these conformations are often critical for protein interactions and complex formation, they are lowly populated, short-lived and cannot be isolated making them essentially invisible to conventional structural biology techniques.⁽⁵⁸⁾

Fortunately, pNMR is highly sensitive to minor states and is therefore ideally suited for studying lowly populated conformations (between 0.5 and 50% of the total population).^(58, 59) This makes pNMR particularly useful for studying ET complexes as they often utilize an intermediate encounter complex to balance their high turn-over rates with the need for specific interactions.⁽⁵⁷⁾ This is the case for the ET complex between Cc and CcP to which pNMR techniques have been successfully applied previously.^(1, 8, 9, 35, 60)

pNMR relies on three main paramagnetic effects, including PCS, PRE and partial alignment resulting in RDC. These effects are generated by the unpaired electrons of paramagnetic centres and affect near-by nuclei in such a way that long-range distance information and/or bond vector orientations can be obtained.⁽⁶¹⁾ The resulting data generally represent the time-averaged set of conformations present in the sample and can be converted into structural restraints for protein docking and ensemble modelling.⁽⁶²⁾

There are several different paramagnetic centres available that are useful for pNMR. Which type of effect a given centre generates depends on two properties of the centre: PRE is governed by the electronic longitudinal relaxation rate and PCS and RDC are governed by the time-averaged anisotropic component of the magnetic susceptibility tensor (χ -tensor). For example, paramagnetic centres with an isotropic χ -tensor and low longitudinal relaxation rate (10^7 - 10^9 s⁻¹), such as the nitroxide radicals of spin labels (SLs) or the metal ions Mn²⁺ and Gd³⁺, generate large PRE but no PCS or RDC. Conversely, centres with anisotropic χ -tensors and a high longitudinal relaxation rate (10^{10} - 10^{13} s⁻¹), such as the metal ions Co²⁺, Fe³⁺ and most lanthanoids, generate primarily PCS and RDC.^(58, 63) A summary of the effects generated by several different paramagnetic centres commonly used in pNMR can be seen in Table 2.1.

TABLE 2.1 Paramagnetic properties of various metals. The columns RDC, PRE and PCS indicate the suitability of the metal to obtain these restraints and the distance range in which significant PRE or PCS are observed is indicated. HS and LS refer to high-spin and low-spin, respectively. Table adapted from reference (62).

Paramagnet	S/J ^a	τ_s^b , s	RDC	PRE	PCS	distance range (Å)	diamagnetic analogue	ref.
Fe ³⁺ HS	5/2	10 ⁻⁹ -10 ⁻¹¹	-	+	+	5-12 ^c	Fe ²⁺ LS	(64)
Fe ³⁺ LS	1/2	10 ⁻¹¹ -10 ⁻¹³	+	+	+	5-17 ^c	Fe ²⁺ LS	(65)
Mn ²⁺	5/2	10 ⁻⁸	-	++	-	16-25	Ca ²⁺ , Mg ²⁺	(66)
Co ²⁺ LS	3/2	10 ⁻⁹ -10 ⁻¹⁰	+	-	+	4-13	Zn ²⁺ , Cd ²⁺	(67)
Ni ²⁺	2/2	10 ⁻¹⁰ -10 ⁻¹²	-	+	+	12-22 ^c	Zn ²⁺ , Cd ²⁺	(68)
Cu ²⁺	1/2	10 ⁻⁸ -10 ⁻⁹	-	+	-	5-25	Cu ⁺ , Zn ²⁺	(69)
Gd ³⁺	7/2	10 ⁻⁸ -10 ⁻⁹	-	++	-	20-30	La ³⁺ , Lu ³⁺ , Y ³⁺	(70)
Er ³⁺	15/2	10 ⁻¹² -10 ⁻¹³	+	-	++	10-40	La ³⁺ , Lu ³⁺ , Y ³⁺	(71)
Ho ³⁺	16/2	10 ⁻¹² -10 ⁻¹³	+	-	++	10-40	La ³⁺ , Lu ³⁺ , Y ³⁺	(72)
Yb ³⁺	7/2	10 ⁻¹² -10 ⁻¹³	+	-	++	10-40	La ³⁺ , Lu ³⁺ , Y ³⁺	(73)
Dy ³⁺	15/2	10 ⁻¹² -10 ⁻¹³	++	-	++	15-60	La ³⁺ , Lu ³⁺ , Y ³⁺	(74)
Tb ³⁺	12/2	10 ⁻¹² -10 ⁻¹³	++	-	++	15-60	La ³⁺ , Lu ³⁺ , Y ³⁺	(75)
Tm ³⁺	12/2	10 ⁻¹² -10 ⁻¹³	++	-	++	15-60	La ³⁺ , Lu ³⁺ , Y ³⁺	(76)
Nitroxide	1/2	10 ⁻⁷	-	++	-	12-25	reduced radical, MTS ^d	(77)

^a J is the quantum number used for lanthanoids, which is a vectorial addition of S and L, the total orbital angular momentum quantum number;⁽⁷⁸⁾ ^b τ_s is the electronic relaxation time which is the inverse of the equivalent rate; ^c based on PRE; ^d MTS, (1-Acetyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-methanethiosulfonate, diamagnetic analogue of MTSL.⁽¹⁾

Although a few paramagnetic centres do occur naturally in proteins, such as the Fe³⁺ in haem groups, they are rare and do not often produce relevant/useful data. Therefore, several methods are available for introducing such centres experimentally onto proteins. Free paramagnetic centres in solution can be used to study surface exclusion during complex formation.^(58, 79, 80) However, site-specific tagging provides much more detailed structural information⁽⁵⁸⁾ and such tags can be attached non-covalently^(81, 82) or via fusion peptides, surface thiols or unnatural amino acids.⁽⁸³⁾ In any case, when designing an experiment, consideration must be given to the distance over which a given centre can generate paramagnetic effects and to the flexibility of the tag, which can cause signal averaging.^(84, 85)

PARAMAGNETIC NMR THEORY

PARAMAGNETIC RELAXATION ENHANCEMENT

To date, the most commonly used paramagnetic effect in structural studies has been PRE. PRE is an observed increase in the relaxation rate for nuclei close to the paramagnetic centre.⁽⁸⁶⁾ PRE can occur via either the Solomon mechanism or the Curie mechanism depending on the electronic relaxation rate of the unpaired electrons in the paramagnetic centre.⁽⁸⁷⁾ The Solomon mechanism, or dipole-dipole relaxation, is dominant for isotropic paramagnetic centres with a long relaxation time such as nitroxide radicals (MTSL [1-oxy-2,2,5,5-tetramethyl-2,5-dihydropyrrol-3-ylmethyl methane thiosulfonate]; pyMTSL [3-methanesulfonylthiomethyl-4-(pyridin-3-yl)-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-1-yloxy]) or some metal ions such as Mn^{2+} and Gd^{3+} .⁽⁵⁸⁾ It affects both the longitudinal ($R_{1,dipole}$) and transverse ($R_{2,dipole}$) relaxation rates according to equations 2.1 and 2.2 (see list of symbols at beginning of thesis for all symbol descriptions):

$$R_{1,dipole} = \frac{2}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_i^2 g_i^2 \beta^2 S(S+1)}{r_{IM}^6} \left(\frac{3\tau_c}{1 + \omega_i^2 \tau_c^2} \right) \quad 2.1$$

$$R_{2,dipole} = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_i^2 g_i^2 \beta^2 S(S+1)}{r_{IM}^6} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_i^2 \tau_c^2} \right) \quad 2.2$$

Since this form of PRE is proportional to γ_I^2 , ^1H nuclei are roughly 100 and 16 fold more sensitive to PRE than ^{15}N or ^{13}C nuclei, respectively.⁽⁶³⁾

For paramagnetic metal ions such as lanthanoids (Ln^{3+}) that have anisotropic χ -tensors, the Curie mechanism is dominant, although it is only dependent on the isotropic components of the χ -tensor.⁽⁸⁸⁾ It affects predominantly the transverse ($R_{2,Curie-spin}$) relaxation rate according to equation 2.3:

$$R_{2,Curie-spin} = \frac{1}{5} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\omega_i^2 g_i^4 \beta^4 S^2(S+1)^2}{(3kT)^2 r_{IM}^6} \left(4\tau_r + \frac{3\tau_r}{1 + \omega_i^2 \tau_r^2} \right) \quad 2.3$$

Since this mechanism is dependent on τ_r and not on τ_s , and it scales with ω_I^2 , it is mainly important for large proteins in a strong magnetic field.⁽⁶³⁾ The effects of the Curie mechanism on longitudinal relaxation ($R_{1,Curie-spin}$) are negligible.⁽⁶²⁾

Both forms of PRE are highly distance dependent (proportional to r_{IM}^{-6}) and are therefore particularly useful for studying lowly populated states of highly dynamic proteins or protein complexes. However, in order to obtain information about a lowly populated or minor state, generally it must experience larger PRE, and therefore be closer to the paramagnetic

centre, than the major state. The exchange between the major and minor states must also be faster than the PRE rate experienced by the minor state.⁽⁸⁶⁾

Although PRE can be measured using the increase in either the transverse (R_2) or longitudinal (R_1) relaxation rates, the former ($R_{2,para}$) can be measured more accurately and is less susceptible to internal motions and cross-relaxation than $R_{1,para}$.⁽⁸⁹⁾ For nuclei near to the paramagnetic centre, a decrease in peak intensity in the spectrum of the paramagnetic sample (I_{para}) is observed compared to the diamagnetic sample (I_{dia}) and this can be used to determine $R_{2,para}$ according to equation 2.4:

$$\frac{I_{para}}{I_{dia}} = \frac{R_{2,dia} e^{(-R_{2,para} t)}}{R_{2,dia} + R_{2,para}} \quad 2.4$$

where $R_{2,dia}$ is the transverse relaxation rate for the diamagnetic sample, measured as half the peak width at half height, and t is the period in which the 1H magnetization experiences transverse relaxation during the pulse sequence.⁽⁷⁷⁾ It is also possible to determine $R_{2,para}$ using a two-time-point method according to equation 2.5:

$$R_{2,para} - R_{2,dia} = \frac{1}{T_b - T_a} \ln \frac{I_{dia}(T_b)I_{para}(T_a)}{I_{dia}(T_a)I_{para}(T_b)} \quad 2.5$$

where T_0 is the initial time point and T_1 is the second time point. This prevents underestimations of $R_{2,para}$ that can occur when using a short repetition delay (the paramagnetic sample recovers faster than the diamagnetic sample due to the effect of PRE on $R_{1,para}$) but also greatly decreases the sensitivity of the measurement.⁽⁹⁰⁾ In practice, both methods often yield similar results.⁽⁵⁸⁾

Once calculated, the $R_{2,para}$ values can be transformed into structural restraints for ensemble modelling. The distance (r_{IM}) between the observed nucleus and the paramagnetic centre can be determined according to the equation 2.6:

$$r_{IM} = \sqrt[6]{\frac{1}{R_{2,para}} \frac{\gamma_I^2 g_s^2 \beta^2 \mu_0^2 (S+1) S}{240\pi^2} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_I^2 \tau_c^2} \right)} \quad 2.6$$

where $\tau_c^{-1} = \tau_r^{-1} + \tau_s^{-1}$ where τ_r is the rotational correlation time of the vector that connects the paramagnetic centre and the nucleus and τ_s is the longitudinal electronic relaxation time.^(63, 77)

PSEUDOCONTACT SHIFTS

Anisotropic paramagnetic centres can cause paramagnetic shifts for the resonances of nearby nuclei. These changes in the resonance positions are the result of both the through-bond Fermi contact shift and the through-space PCS.⁽⁹¹⁾ Fermi contact shifts are generally used to study ligands that are bound directly to the paramagnetic centre as the magnitude of the shift decays rapidly with the number of chemical bonds between the nucleus and the paramagnetic centre.⁽⁹²⁾ On the other hand, PCS can provide structure and orientation information over long distances, up to 70 Å,^(93, 94) and are highly predictable making them easy to interpret.⁽⁹⁵⁾ Chemical shift anisotropy can also contribute to the shift difference but this contribution is small and can be corrected for during data analysis.⁽⁹⁶⁾

PCS can be determined experimentally by comparing the resonance positions observed for a paramagnetic sample to those for a diamagnetic sample. The shift difference (δ_{PCS}) can then be converted into modelling restraints using equation 2.7:

$$\delta_{PCS} = \frac{1}{12\pi r_{IM}^3} \left(\Delta\chi_{ax} (3\cos^2\theta - 1) + \frac{3}{2} \Delta\chi_{rh} (\sin^2\theta \cos 2\varphi) \right) \quad 2.7$$

where r_{IM} and the angles θ and φ provide the polar coordinates of a given nucleus within the frame of the χ -tensor, which is characterized by $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ and originates at the paramagnetic centre.⁽⁶³⁾ Starting with as few as twenty measured PCS, the χ -tensor and remaining PCS can be predicted for use in further resonance assignment.^(96, 97)

RESIDUAL DIPOLAR COUPLING

Anisotropic paramagnetic centres will align themselves with an external magnetic field, such as the B_0 field in NMR, and consequently induce partial alignment of any molecules to which they are bound. When this occurs, anisotropic effects no longer average to zero and small residual effects can be observed such as RDC.⁽⁶¹⁾ RDC provides information about the orientation of bond-vectors with respect to the χ -tensor and so can be used to determine the orientation of protein domains or whole proteins within complexes.⁽⁹⁸⁾ A convenient feature of RDC is that the effects are independent of the distance between the nucleus and the paramagnetic center.⁽⁹⁹⁾

RDC is observed as an increase in the J-coupling between nuclei and so can be measured by comparing the J-coupling observed for a paramagnetic sample to that seen for a

diamagnetic sample.^(30, 99) The RDC (D^{res}) can be converted into modeling restraints using equation 2.8:⁽⁹⁹⁾

$$D^{\text{res}} = -\frac{B_0^2}{15kT} \frac{\gamma_A \gamma_B \hbar}{16\pi^3 r_{AB}^3} \left(\Delta\chi_{ax} (3\cos^2 \Theta - 1) + \frac{3}{2} \Delta\chi_{rh} (\sin^2 \Theta \cos 2\phi) \right) \quad 2.8$$

While the χ -tensor is degenerate and multiple RDC values can be fit to the same Θ and ϕ values,⁽³²⁾ this problem is easily solved by combining data sets from multiple lanthanoids,^(71, 81, 100, 101) from paramagnetic centres at different positions on the protein^(102, 103) or from different paramagnetic effects like PRE^(80, 104-108) or PCS.^(101, 109-112)

PROTEINS USED IN THIS STUDY

CYTOCHROME C

Cc is a small, haem containing protein that is loosely associated with the inner mitochondrial membrane where it is an electron transporter in the eukaryotic respiratory chain. It was discovered in *S. cerevisiae* in 1925 by D. Keilin after he noticed unexpected spectrophotometric properties in fly larvae, bacteria and yeast cells.⁽¹¹³⁻¹¹⁵⁾ Since then, the sequences from hundreds of different species have been reported and they show that Cc is highly conserved among eukaryotes.⁽¹¹⁶⁾ The protein sequence consists of a core of approximately 104 amino acids plus an additional four to nine amino acids on the N-terminus for some organisms (Figure 2.1).

Organism	Residue number						
	-9	1	10	20	30	40	50
<i>Arabidopsis</i>	-QVAD ISLQ	GDAKKGANLF	KTRCAQC HTL	KAGEGNKIGF	ELHGLFGRKT	GSVAGYS YTD	
<i>C. elegans</i>	----SDIPA	GDYEKGKKVY	KQRCLQC HVV	DS-TATKTGP	TLHGVIGRKS	GTVSGFDYSA	
Fruit fly <i>iso-1</i>	-----GS	GDAENGGKIF	VQKCAQC HTY	EVGGKHKVGP	NLGGVVGRC	GTAAGYK YTD	
Fruit fly <i>iso-2</i>	-----GVPA	GDVEKGKKLF	VQKCAQC HTV	EAGGKHKVGP	NLHGLIGRKT	GQAAGFA YTD	
Horse	-----	GDVEKGKKIF	VQKCAQC HTV	ERGGKHKTGP	NLHGLFGRKT	GQAPGPTYTD	
Human	-----	GDVEKGKKIF	IMKCSQC HTV	ERGGKHKTGP	NLHGLFGRKT	GQAPGYS YTA	
Mouse	-----	GDVEKGKKIF	VQKCAQC HTV	ERGGKHKTGP	NLHGLFGRKT	GQAAGFS YTD	
Rice	-ASFSEAPP	GNPKAGEKIF	KTKCAQC HTV	DKGAGHKQGP	NLNLGFRQS	GTPPGYS YST	
Wheat	-ASFSEAPP	GMPDAGAKIF	KTKCAQC HTV	DAGAGHKQGP	NLHGLFGRQS	GTTAGYS YSA	
Yeast <i>iso-1</i>	----TEFKA	GSAKKGATLF	KTRCLQC HTV	EKGGPHKVGPF	NLHGIFGRHS	GQAEGYS YTD	
Yeast <i>iso-2</i>	AKESTGFKP	GSAKKGATLF	KTRCQC HTI	EEGGPNKVGPF	NLHGIFGRHS	GQVKGYS YTD	
Zebra fish	-----	GDVEKGKKVF	VQKCAQC HTV	ENGGKHKVGP	NLWGLFGRKT	GQAEGFS YTD	
Clustal analysis		* . . * . : :	: * * * * .	. * * * *	* * * * * : * * *	* . * : * *	* . * : *

Organism	Residue number				
	60	70	80	90	100
<i>Arabidopsis</i>	ANKQKGI EWK	DDTLFEY LEN	PKKYIPG TKM	AFGGLK KPKD	RNDLIT FL EE ET K---
<i>C. elegans</i>	ANKNKG VVWT	KETLFEY LLN	PKKYIPG TKM	VFAGL KKADE	RADLI KY IEV ES AKSL
Fruit fly <i>iso-1</i>	ANIKKG V TWT	EGNLDEY LKD	PKKYIPG TKM	VFAGL KKA EE	RADLI AF LKS NK ----
Fruit fly <i>iso-2</i>	ANKAKG I TWN	EDTLFEY LEN	PKKYIPG TKM	IFAGL KKP NE	RGDLI AY LKS ATK ----
Horse	ANKNKG I TWK	EETLMEY LEN	PKKYIPG TKM	IFAGI KK TE	REDLI AY LK AT NE--
Human	ANKNKG I IWG	EDTLMEY LEN	PKKYIPG TKM	IFVGI KK EE	RADLI AY LK AT NE--
Mouse	ANKNKG I TWG	EDTLMEY LEN	PKKYIPG TKM	IFAGI KK GE	RADLI AY LK AT NE--
Rice	ANKNMA VI WE	ENTLYD YLLN	PKKYIPG TKM	VFPGL KK PQE	RADLI S YLKE AT S---
Wheat	ANKNK A VEWE	ENTLYD YLLN	PKKYIPG TKM	VFPGL KK PQD	RADLI AY LK AT S--
Yeast <i>iso-1</i>	ANIKKN V LWD	ENNMSEY L TN	PKKYIPG TKM	AFGGL KK EKD	RNDLI T YLK ACE---
Yeast <i>iso-2</i>	ANINKN V KWD	EDSMSEY L TN	PKKYIPG TKM	AFAGL KK EKD	RNDLI T Y MTK AAK ---
Zebra fish	ANKSKG I VWG	EDTLMEY LEN	PKKYIPG TKM	IFAGI KK GE	RADLI AY LK AT S---
Clustal analysis	** : *	. . : : ** :	*****	* * : * * :	* * * * : :

FIGURE 2.1 Multiple sequence alignment for Cc from several eukaryotes. Haem iron ligands are shown in bold. Fully conserved residues are marked with a star (*), residues with strongly similar properties are marked with a colon (:), and residues with weakly similar properties are marked with a period (.). Sequences were obtained from UniProt⁽¹¹⁷⁾ and alignment was done with Clustal Omega.^(118, 119)

Due to the high level of conservation between species in the primary protein sequence and the overall tertiary structure, Cc from several other eukaryotic species, in particular horse, have often been used as structural and functional mimics of yeast Cc; however, it must be noted that there are some observable differences in the way they interact with CcP compared to yeast Cc.⁽¹²⁰⁾

Of the two isoforms found in yeast, *iso-1* is most commonly studied⁽¹²⁰⁾ and was also used in this work. Native *iso-1* Cc can be easily purified directly from *S. cerevisiae*⁽¹²¹⁻¹²³⁾ and includes the posttranslational tri-methylation of K72.⁽¹²⁰⁾ Alternatively, recombinant protein can be expressed and purified in *E. coli*.^(124, 125) In this system, the genes for both yeast *iso-1* Cc (CYC1) and Cc haem lyase (CYC3), which catalyzes the covalent insertion of the haem group in the cytoplasm,⁽¹²⁶⁾ are co-expressed, resulting in good yields of active *holo*-protein (~20 mg/L culture for rich media). Cc has a very high solubility of ~100 g/L in buffer.

Yeast Cc consists of 108 amino acids resulting in a 12.7 kDa positively charged protein with a pI of approximately 9.54.⁽¹²⁰⁾ The highly conserved protein fold consists of five α -helices and a short β -strand that essentially form a sphere (Figure 2.2A).⁽¹¹⁶⁾

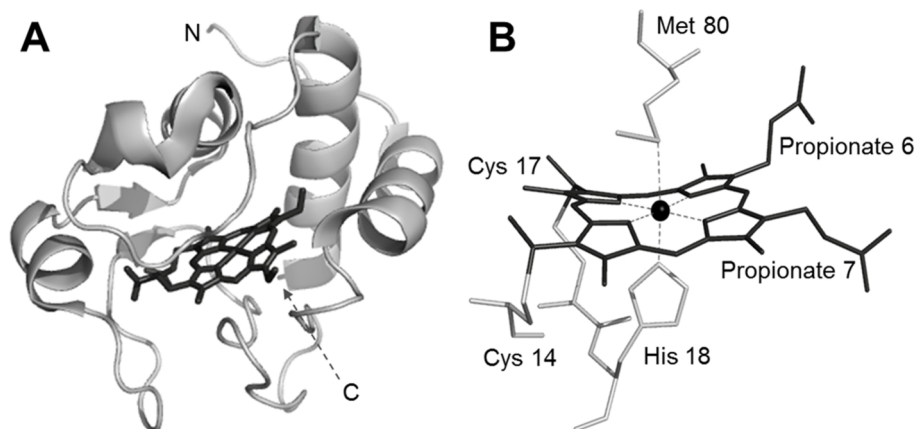


FIGURE 2.2 Three dimensional structure of yeast *iso*-1 Cc. **A)** Overall view of the protein structure (show in ribbon) showing five α -helices and one β -sheet with the N- and C-termini labelled. The haem group is shown as sticks. **B)** The haem binding site showing the iron ion (black sphere) coordinated to four pyrrole nitrogens and the two axial ligands, M80 and H18. The haem group is covalently attached to the protein via C14 and C17 and the two propionate groups are also indicated. The molecular coordinates are for ferric Cc (PDB entry 2YCC).⁽¹²⁷⁾

At the centre of the protein is a covalently bound (*c*-type) haem group that is attached via thioether bonds to two cysteine residues in the sequence CXXCH (residues 14-18 in Figure 2.1) (Figure 2.2B). The haem group contains a LS ($S=1/2$), hexacoordinate iron ion that is equatorially coordinated to the four pyrrole nitrogens of haem itself and axially coordinated to the side chains of H18 (N_2) and M80 (S_8) from Cc. During ET, the iron interconverts between reduced Fe^{2+} , which is diamagnetic, and oxidized Fe^{3+} , which is paramagnetic.⁽¹¹⁶⁾ Many high-resolution structures, determined by both X-ray crystallography and solution NMR spectroscopy, are available for the ferrous^(128, 129) and ferric forms.^(127, 130, 131)

The ability of the haem iron to rapidly interconvert between a ferrous and ferric state is key to the physiological role of Cc as an electron transporter in cellular oxidation.⁽¹³²⁾ In addition to its standard function as an electron transporter between the membrane bound Cc reductase (complex III) and oxidase (complex IV) during oxidative phosphorylation, yeast Cc also has two additional physiological partners: lactate dehydrogenase (or cyt *b*₂) and CcP,⁽¹¹⁶⁾ the latter of which will be discussed below. Furthermore, Cc also plays a key role in the regulation of apoptosis in response to DNA damage.⁽¹³³⁾

CYTOCHROME C PEROXIDASE

CcP is a water soluble haem protein that belongs to the peroxidase family and was discovered in *S. cerevisiae* in 1940 by Altschul, Abrams and Hogness.⁽¹³⁴⁾ It is found in the mitochondrial intermembrane space where it catalyzes the reduction of potentially harmful hydrogen peroxide to water using electrons received from its physiological partner, Cc. CcP can also react with other hydroperoxides but the reaction rate is much slower.⁽¹³⁵⁾

Yeast CcP consists of 294 amino acids resulting in a 34 kDa, negatively charged protein with a pI of 4.5. Native CcP can be purified directly from *S. cerevisiae*⁽¹³⁵⁻¹³⁷⁾ or recombinant protein can be expressed and purified from *E. coli* using one of several different expression systems.⁽¹³⁸⁻¹⁴³⁾ Although bacterial expression produces the *apo*-protein and a non-covalent *b*-type haem group must be added during purification, this process is trivial and compensated for by very high yields, especially for isotopically labelled protein (~130 mg/L culture for minimal media in D₂O).⁽¹⁴⁰⁾

Interestingly, CcP crystallizes when dialyzed against pure water and it was therefore the first haem enzyme for which the X-ray crystal structure was solved (Figure 2.3A).⁽¹⁴⁴⁾

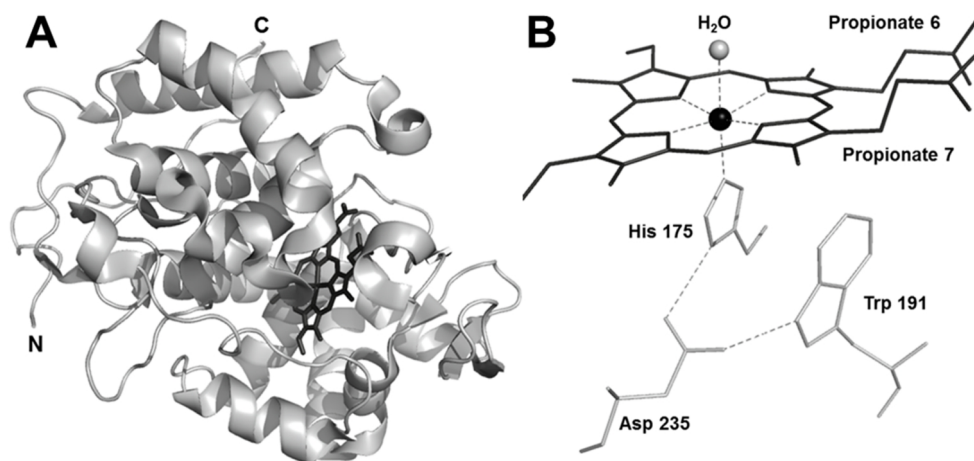


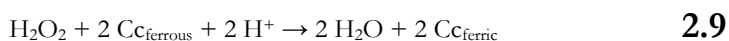
FIGURE 2.3 Three dimensional structure of yeast CcP. **A)** Overall view of the protein structure (shown in ribbon) showing the dominant α -helical structure. The N- and C-termini are labelled and the haem group is shown as sticks. **B)** The haem binding site showing the iron ion (black sphere) coordinated to four pyrrole nitrogens and the two axial ligands, H175 and water. The two propionate groups and residues D235 and W191 are also indicated. The molecular coordinates are for resting state CcP (PDB entry 1ZBY).⁽¹⁴⁴⁾

The secondary structure is dominated by α -helices and with a *b*-type haem at the center, buried within a highly conserved hydrophobic pocket. The iron of the haem group is equatorially coordinated to the four pyrrole nitrogens of haem itself and axially coordinated to the side chain of H175 (N_{e2}) (Figure 2.3B); the sixth coordination position is where hydrogen peroxide binds during catalysis.⁽¹⁴⁴⁾ In the resting state (RS), the haem group contains a paramagnetic, pentacoordinate high-spin (HS) ($S=5/2$) Fe^{3+} atom. During the formation of compound I (CpdI), the first step of catalysis, the iron is oxidized to Fe^{4+} and forms an oxyferryl intermediate.^(135, 145) Depending on experimental conditions, other molecules can be coordinated in the sixth position resulting in a hexacoordinate low-spin (LS) ($S=1/2$) iron ion (for strong ligands according to ligand field theory like CN^- or NO_2) or a hexacoordinate HS ($S=5/2$) iron ion (for weak ligands according to ligand field theory like H_2O or Cl^-); both of which are paramagnetic.⁽¹⁴⁶⁾ Several high resolution crystal structures are available of the resting enzyme and CpdI.⁽¹⁴⁷⁻¹⁴⁹⁾

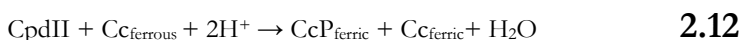
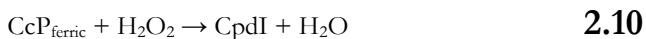
Although CcP is not directly involved in metabolic respiration, its ability to rapidly accept electrons from Cc makes it a useful structural model for other proteins of the electron transport chain. Furthermore, the reaction mechanism of CcP is very similar to that of other members of the peroxidase family and thus it is widely used as a model when studying the reduction of hydroperoxides.⁽¹³⁵⁾

THE CC-CCP COMPLEX

The Cc-CcP complex has been widely studied as a prototypical ET complex and serves as a model for the more complex reactions that occur in the respiratory electron transport chain.⁽⁶⁾ CcP catalyzes the reduction of hydrogen peroxide to water according to equation 2.9:



CcP reduces hydrogen peroxide using two reducing equivalents that it receives from ferrous Cc, producing water and ferric Cc.⁽¹³⁵⁾ This reaction occurs in at least three steps according to equations 2.10-2.12:



First, ferric CcP reacts with hydrogen peroxide to form water and the intermediate CpdI, in which the haem iron is oxidized to an oxy-ferryl group^(150, 151) and the side chain of W191 is oxidized to a cationic indole radical (equation 2.10).⁽¹⁵²⁻¹⁵⁹⁾ The aromatic ring of W191 is stabilized by vdW interactions with the imidazole ring of H175, and both are in turn stabilized by hydrogen bonds to D235 (Figure 2.3B).⁽¹⁴⁴⁾ In the next step, CcP forms a complex with ferrous Cc and CpdI is reduced to compound II (CpdII) via the indole radical (equation 2.11).⁽¹⁶⁰⁻¹⁶⁸⁾ There is a fast equilibrium between the indole radical and the oxy-ferryl so that the radical is quickly regenerated before being reduced by a second round of ET.^(163, 164) In the final step, a second ferrous Cc binds and reduces CpdII, thus returning CcP to the RS (equation 2.12).^(161, 164)

The complex between Cc and CcP is one of the few highly dynamic complexes that has been crystallized (Figure 2.4).⁽¹⁶⁹⁾

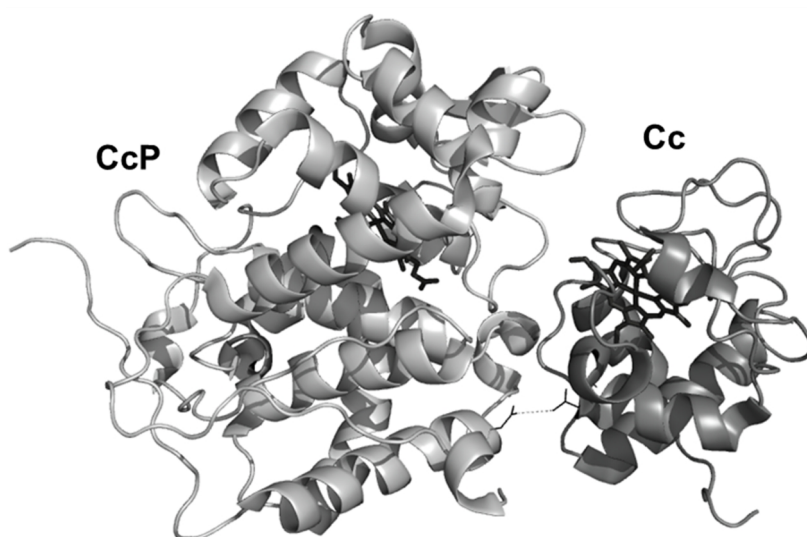


FIGURE 2.4 Three dimensional crystal structure of the complex between yeast Cc (dark grey) and yeast CcP (light grey). The complex is stabilized by a single hydrogen bond (dotted line) between CcP E290 and Cc N70 at the bottom of the structure (shown as sticks). The protein backbones are shown as ribbons and the haem groups are shown in black sticks (PDB entry 2PCC).⁽¹⁶⁹⁾

Unexpectedly for such a large complex, the crystal structure showed that the complex is stabilized by a single hydrogen bond between N70 of Cc and E290 of CcP.⁽¹⁶⁹⁾ However, a recent study suggested possibly two additional hydrogen bonds between Q116-A193 and K87-E32 (Cc-CcP)⁽¹⁷⁰⁾ and it is also thought that further hydrogen bonds to ordered water molecules within the interface are present.⁽¹⁶⁹⁾ In contrast, the crystallized complex between

CcP and horse Cc has several potential hydrogen bonds between K72-E290, K8-N38, K87-E35⁽¹⁶⁹⁾ K13-D34, E90-D34 and Q12-N195 (Cc-CcP).⁽¹⁷⁰⁾ However, both complexes are mainly stabilized by electrostatic interactions or vdW forces between specific residues.⁽¹⁶⁹⁾

The interaction between Cc and CcP has been extensively studied using site-specific mutagenesis. On CcP, several residues within or near the binding site have been shown to be important for complex formation: R31, E35, D34, E35, D37, E98, E118, A193, V197, E201, E209, E290.^(162, 163, 171-180) In particular, V197 of CcP has been shown to form multiple vdW interactions with A81 of Cc ^(169, 172) and the same holds for Y39 of CcP and R13 of Cc.^(60, 169) Furthermore, several charged residues of CcP that are further away from the binding site do not play a role in complex formation: E32, D79, D217 and E291.^(162, 163, 173-177) These results are agreement with X-ray crystallography work⁽¹⁶⁹⁾ and similar work done on the complex with horse Cc.^(162, 163, 171, 173-175, 177, 181)

Similar studies have identified several Cc residues that are critical for complex formation including K72, K87⁽¹⁷²⁾ and particularly F82;⁽¹⁸²⁾ This is in line with the data from X-ray crystallography⁽¹⁶⁹⁾ and NMR spectroscopy.^(183, 184) Residue F82 has also been shown to be important for intermolecular ET.^(185, 186) In contrast, the highly conserved residue K27 apparently plays no role in the interaction.⁽¹⁸⁷⁾ Another interesting result is that the mutations D34A⁽¹⁷²⁾ and T12A,⁽⁶⁰⁾ located in the binding interfaces of CcP and Cc respectively, actually result in tighter complex formation. For T12A, the structural causes for this are unknown. However, for D34, the mutation from a negative to a neutral residue could lessen repulsion between D34 and nearby D90 on Cc allowing for tighter complex formation.⁽¹⁴⁶⁾

The stoichiometry of the complex has been a matter of debate for many decades. In the crystal structure⁽¹⁶⁹⁾ and under many experimental conditions, a 1:1 stoichiometry has been observed.⁽¹²⁰⁾ However, a second low-affinity binding site (two to four orders of magnitude weaker binding) has been proposed.⁽¹⁸⁸⁻¹⁹⁰⁾ Several kinetic studies have supported this, suggesting that there is both a high-affinity, low-reactivity binding site and a low-affinity, high-reactivity binding site, and that Cc binds to both sites simultaneously under low ionic strength conditions (<50 mM salt).^(181, 191-196) BD simulations⁽³⁸⁾ and mutagenesis studies⁽¹⁹⁷⁾ have suggested the area around residue D148 as a possible location for this second binding site. This site and two others near D217 and Y39 were suggested by simulations of electrostatic interactions.^(198, 199) In 2015, a structural pNMR study was done in which Cc was cross-linked to CcP at the stereo-specific binding site. EDTA-Mn²⁺ tags were attached to free Cc and the PRE effects were then observed on CcP-Cc cross-linked complex. A low-affinity complex was found consisting of an ensemble of minor states covering two spatial regions: one close to CcP

residues D148 and D217 and extending to D33 and E35 and the other close to CcP residues E167, D261 and E267. Charge reversal mutations confirmed the importance of CcP residues D148 and D217 for low-affinity complex formation.⁽²⁰⁰⁾

This debate has spawned several hypotheses for how the complex is formed as well as for the mechanism of intermolecular ET. As mentioned above, one hypothesis for complex formation postulates a two-binding-site model; a high-affinity, low-reactivity binding site as seen in the crystal structure and a low-affinity, high-reactivity binding site elsewhere with no interaction between the two bound Cc molecules.⁽¹²⁰⁾ Evidence for this model comes predominantly from kinetic studies^(181, 188, 190-195, 197, 201) and computer simulations.^(38, 198, 199) An alternative hypothesis proposes that two Cc molecules cannot bind simultaneously due to strong electrostatic repulsion between them; instead several versions of the singly bound complex exist with Cc bound at different locations, only a sub-set of which are represented by the crystal structure.^(38, 198) This model matches well solution NMR studies that have shown that the complex is highly dynamic.^(1, 8, 184, 202, 203) In solution, the complex exists partially as a weakly bound encounter complex, in which Cc samples ~15% of the CcP surface. However, the orientation observed in the crystal structure is indeed the dominant form; it accounts for 70% of the lifetime of the complex while the remaining 30% of the time is spent in the encounter complex.^(1, 8, 35) Furthermore, these lifetimes seem to be the same regardless of the redox state of the two proteins.⁽²⁰⁴⁾ Intriguingly, the percentage of time spent in the encounter complex can be modified to as little as 10% or as much as 80% by modulating affinity via binding interface point mutations.⁽⁹⁾ It must be noted however that these NMR studies are done under moderate ionic strength conditions (≥ 100 mM) as opposed to the low ionic strength conditions of the studies in which the second binding site has been observed.

Similar to complex formation, multiple views for the mechanism of intermolecular ET also exist. The first model is related to the two-binding-site model mentioned above, but posits that only the high-affinity binding site is ET active. In this model, binding of Cc at the low-affinity binding site under low ionic strength works to promote dissociation of Cc from the high-affinity binding site.^(175, 201, 205, 206) This view is supported by ET kinetics,⁽¹⁶⁸⁾ steady-state kinetics,^(176, 181) stopped-flow spectroscopy⁽²⁰⁷⁾ and cross-linking studies.⁽²⁰⁸⁻²¹¹⁾ The opposing model, which corresponds to the second hypothesis for complex formation mentioned above, suggests that a single Cc can bind at multiple locations on the surface of CcP and then electrons can be transferred via either site, although the transfer rate is much faster via the low-affinity site.^(192, 193) The evidence for this hypothesis has been obtained mainly using photo-chemistry techniques with the complex both in solution^(192-194, 196, 212) and in

the crystal state.^(182, 213) Although the distances between the haem groups indicate that haem-haem ET is only possible in the low-affinity complex,⁽²⁰⁰⁾ no enzymatic turnover is observed when the high-affinity site is blocked but haem-haem ET via the low affinity site remains possible.⁽²⁰⁸⁻²¹¹⁾ Furthermore, the low-affinity complex does not form at physiological salt concentrations suggesting that the biological enzymatic reaction involves ET via W191 at the high-affinity binding site.⁽²⁰⁰⁾ For more detailed information on the Cc-CcP complex, the reader is referred to the recent reviews in references ⁽¹²⁰⁾ and ⁽²¹⁴⁾.

THESIS OUTLINE

The aim of the research presented in this thesis was to build upon the work of previous studies and further characterize the interaction between Cc and CcP, with a particular emphasis on the encounter complex. This was accomplished with the use of site-directed mutagenesis in combination with paramagnetic tags and pNMR spectroscopy.

Chapter I provides a brief overview of the current methods for studying and visualizing encounter complexes and illustrates these using several recent examples from ET complexes.

Chapter II gives an introduction to the theory and use of pNMR as well as an introduction to the proteins used in this work, Cc and CcP.

Chapter III describes the optimization of the CcP expression and purification protocol followed by the use of 3D NMR spectroscopy in order to assign the protein amide backbone of CcP. This is followed by the characterization of the haem iron $\Delta\chi$ -tensors for two forms of yeast CcP: the resting state and HF bound CcP.

Chapter IV describes the use of site specific spin labelling of Cc to observe the PRE effects on CcP and compare those results to previous studies done in the opposite manner, which highlights the importance of obtaining multiple data sets when studying encounter complexes.

Chapter V describes the use of intra- molecular PRE data to determine the most favourable SL orientations prior to inter- molecular PRE based protein docking. Although this can improve the accuracy of the docking, this study also highlights the limitations of PRE for determining protein orientations within a complex.

Chapter VI describes ultra-weak inter-molecular interactions observed between CcP molecules in solution. This study demonstrates the exquisite sensitivity of PRE for minor states.

Chapter VII describes the combination of several paramagnetic effects (PRE, PCS and RDC) to study the Cc-CcP complex. Caged lanthanoid NMR probes (CLaNP) on the surface of CcP were used to generate both inter- and intra- molecular effects and the results are compared to previous PRE based models highlighting the advantages of combining different pNMR data sets.

Chapter VII provides a general discussion of this work in relation to other recent studies involving the encounter complex and pNMR.

CHAPTER 3

NMR assignment and analysis of haem iron $\Delta\chi$ -tensors for two forms of yeast cytochrome *c* peroxidase: serendipitous discovery of time-dependent hydrogen fluoride binding

ABSTRACT

The ET complex between yeast Cc-CcP is a well-studied model for highly dynamic encounter complexes. Both the crystal structure and the solution structure of the complex have been available for many years. However, the solution NMR studies have focused on the spectrum of Cc since no backbone resonance assignment for CcP was available due to difficulties with protein solubility and stability. In this study, a novel expression and purification protocol is described that regularly produces high yields of CcP (both natural abundance and isotope enriched) and the protein is highly soluble and has enough thermo-stability to undergo several days of NMR measurements. Furthermore, during these long term NMR experiments, a second form of the protein was detected. After analysis of the paramagnetic properties of both states, it appears that both forms contain haem groups with a hexacoordinate HS iron. The minor form appears to be due to HF binding as the sample ages.

INTRODUCTION

The Cc-CcP complex has been widely studied as a prototypical ET complex and serves as a model for the more complex reactions that occur in the respiratory electron transport chain.⁽⁶⁾ CcP is a 34 kDa haem protein from *Saccharomyces cerevisiae* (Baker's yeast). It is found in the intermembrane space of the mitochondria where it catalyzes the reduction of hydrogen peroxide to water using electrons it receives from its physiological partner, Cc.⁽¹³⁵⁾ This ET occurs via a non-covalent *b*-type haem group that contains a paramagnetic, pentacoordinate HS ($S=5/2$) Fe^{3+} atom at rest. During the formation of CpdI, the first step of catalysis, the iron is oxidized to an oxyferryl (Fe^{+5}) intermediate ($S=1$)^(135, 145) and the side chain of W191 is oxidized to a cationic indole radical.⁽¹⁵²⁻¹⁵⁹⁾ During the second step in catalysis, CcP forms a complex with ferrous Cc and CpdI is reduced to CpdII.⁽¹⁶⁰⁻¹⁶⁸⁾ A second ferrous Cc then reacts with CcP to return CcP to the RS, completing the catalytic cycle.^(161, 164)

This complex is one of the few highly dynamic complexes to have been crystallized and the resulting crystal structure has provided a detailed view of the stereo-specific complex.⁽¹⁶⁹⁾ Additionally, NMR studies have shown that in solution the complex consists of multiple conformations comprising a weakly interacting encounter complex ($\sim 30\%$ of the lifetime of the complex) as well as the stereo-specific state ($\sim 70\%$ of the lifetime of the complex).^(1, 8, 35, 184, 202, 203) However, in all of these studies, information about the complex was obtained by observing only Cc. In order to achieve a more complete understanding of the complex in solution, data should also be obtained for the CcP residues; but, this would require a backbone amide resonance assignment.

Although CcP has been well studied for over 75 years,⁽¹³⁴⁾ a backbone amide resonance assignment was been delayed by low solubility and poor thermo-stability, especially at low ionic strengths.⁽²¹⁵⁾ Recently, the backbone assignment was published by two groups independently^(140, 141, 216) and the first studies characterizing the complex using NMR data from CcP were reported (Chapter 4).^(140, 217-219) In this study, we report a serendipitous and curious observation made during the backbone assignment of CcP (Chapter 3)⁽¹⁴⁰⁾ in which two paramagnetic forms of the protein were observed in the spectra.

Like other haem proteins, the haem group in CcP can exist in several states depending on the nature of the axial ligand bound. This occurs naturally during the steps in the enzymatic cycle but the preference for different ligands can also be influenced by environmental conditions as well as the age of the enzyme.⁽¹⁴⁶⁾ For example, freshly prepared enzyme is generally considered to be pentacoordinate HS (RS CcP).^(146, 220) Depending on

experimental conditions, various ligands can bind resulting in either a hexacoordinate LS ($S=1/2$) iron ion (for strong ligands according to ligand field theory such as CN^- or CO) or a hexacoordinate HS ($S=5/2$) iron ion (for weak ligands according to ligand field theory like H_2O or F^-).⁽¹⁴⁶⁾ This is important when choosing a buffer, especially for pH dependent experiments, as common ions like acetate or chloride can also bind to the haem.^(146, 220-224)

Furthermore, altering the pH can also shift the haem coordination between three states: an unreactive pentacoordinate HS low pH form (below pH 5.5),^(220, 224, 225) the normal active, RS of the enzyme (pentacoordinate HS) at neutral pH and an unreactive hexacoordinate LS high pH form (above pH 8).⁽²²⁶⁻²²⁹⁾ The low pH form is likely caused by a change in the protonation state of either His-52 or the propionic acid side groups of the haem, all of which have pKa values around 6.^(225, 230, 231) The high pH form is a result of an internal coordination of His52 via a water molecule to the haem iron, which can also be caused by aging and/or freezing of the enzyme.^(146, 226, 232-236) Because the reactivity towards hydrogen peroxide varies greatly for these different states, it is important to ensure that the desired state is present in a given sample so that the experimental results are valid.

In this work we have characterized the paramagnetic properties of the haem groups of both the major and minor forms of CcP that we observed during NMR experiments. Both forms appear to contain a hexacoordinate HS iron ion and thus the difference is due to different axial ligands bound to the haem group. The major form is coordinated to water while the minor form appears to be coordinated to HF; previous work has shown that hydrofluoric acid is the dominant ligand at pH 6 and the fluoride anion only becomes a significant ligand above pH 8.⁽²³¹⁾ Therefore, this work also contains the NMR assignment of the amide resonances of HF bound CcP.

MATERIALS & METHODS

SUB-CLONING OF YEAST CCP

The gene construct for *S. cerevisiae* CcP C128A⁽¹⁾ was sub-cloned into a pET28a(+) vector (Novagen, Darmstadt, Germany). The gene was amplified using PCR with a 5' primer containing a *PciI* site (resulting in MSKT as the first four amino acids) and a 3' primer containing an *XbaI* site. The fragment was cloned into a pET28a(+) vector cut with *XbaI* and *NcoI*, which are compatible with *PciI*, yielding pET28aCcP. The sequence of the insertion was verified by DNA sequencing.

EXPRESSION AND PURIFICATION OF CCP

The pET28aCcP plasmid was used to express and purify CcP in a protocol adapted from references^(139, 142) with changes for labelled protein expression and the use of phosphate buffers. The pET28aCcP plasmid was transformed into electrocompetent *E. coli* BL21 Star (DE3)pLysS (Life Technologies Europe BV, Bleiswijk, Netherlands). Several LB pre-cultures containing 0.1% glycerol, 50 µg/mL kanamycin (KAN) and 34 µg/mL chloramphenicol (CAM) were inoculated with a single colony and incubated at 37 °C with shaking at 250 rpm overnight. For rich medium (LB with 0.1% glycerol, 50 µg/mL KAN and 34 µg/mL CAM), 1 mL pre-culture was used to inoculate 750 mL cultures in 2 L Erlenmeyer flasks. For double (¹⁵N,²H) or triple isotope labelled samples (¹⁵N,¹³C,²H), each 2 L Erlenmeyer flask contained 0.5 L of minimal medium (Salts: 13.0 g/L KH₂PO₄, 3.8 g/L K₂HPO₄, 9.0 g/L Na₂HPO₄, 2.4 g/L Na₂SO₄; Trace elements; 32.8 mg/L FeSO₄, 45.3 mg/L CaCl₂, MnCl₂ 7.6 mg/L MnCl₂, 4.4 mg/L CoCl₂, 3.9 mg/L ZnSO₄, 2.3 mg/L CuCl₂, 0.2 mg/L H₃BO₃, 0.050 g/L EDTA; Vitamin supplements: 1.0 mg/L biotin, 1.0 mg/L choline chloride, 1.0 mg/L folic acid, 1.0 mg/L niacinamide, 1.0 mg/L D-pantothenic acid, 1.0 mg/L pyridoxal, 0.1 mg/L riboflavin; and 0.94 g/L MgCl, 30 mg/L thiamine, 1 g/L ¹⁵NH₄Cl (CortecNet, Voisins-Le Bretonneux, France) as the sole nitrogen source prepared in 99.8% D₂O (CortecNet). For triple isotope labelled samples (¹⁵N¹³C²H), 2.0 g/L ¹³C₆-glucose (Sigma Aldrich, Zwijndrecht, Netherlands) was used. The cultures were incubated at 37 °C with shaking at 250 rpm until OD₆₀₀ = 0.6-0.8 and then expression was induced with 0.5 mM IPTG. After induction, rich medium cultures were incubated at 37 °C with shaking at 250 rpm, overnight, double-labelled samples were incubated at 22 °C with shaking at 250 rpm for 24 hours and triple-labelled samples were incubated at 22 °C with shaking at 250 rpm for 72 hours. Cells were harvested by centrifugation at 4 °C at 11,325 x g for 10 minutes. The pellet was re-suspended in approximately 10 mL of cold lysis buffer (100 mM KP_i, 1 mM EDTA, 1 mM PMSF, pH 6.0) and flash frozen in liquid nitrogen and stored at -80 °C.

To purify the protein, the cell pellet was thawed at 4 °C and a few milligrams of lysozyme and 3.8 mg DNAaseI were added. The cells were broken at 120 MPa in a pressure cell homogenizer (FPG12800, Stansted Fluid Power Ltd., Harlow, U.K.). The solution was centrifuged for 45 minutes at 4 °C at 13880 x g to remove cell debris. 30-40 mg of hemin (chloroproporphyrin IX iron(III); Acros Organics, Geel, Belgium) was dissolved in 10 mL lysis buffer with 100 mM NaOH and was added to the supernatant and stirred at 4 °C for 1 hour. The pH of the solution was lowered to 5.0 with a very small volume of 0.5 M HCl and then frozen at -20 °C.

The sample was thawed at 37 °C with shaking at 250 rpm for 2 hours and then centrifuged at 13880 x g at 4 °C for 30 minutes to remove excess hemin. The supernatant was passed through a 0.2 μm polyethersulfon syringe tip filter and was then loaded at 4 mL/min onto a DEAE column (16 cm x 30 cm; DEAE Sephaorse Fast Flow, GE Healthcare, Hoevelaken, Netherlands) that had been equilibrated with 50 mM KPi, pH 5.0. Once the sample was loaded, the column was washed at 4 mL/min with 120 mM KPi until the absorbance at 280 nm fell to baseline. The protein was then eluted with a gradient from 120-500 mM KPi over 40 mL and 5 mL fractions were collected. The fractions containing absorbance at 410 nm were pooled and concentrated to 4 mL using Centriprep centrifugal filter devices (Millipore, Amsterdam Zuidoost, Netherlands) and stored at 4 °C overnight. The sample was then loaded at 1 mL/min onto a gel filtration column (16 cm x 60 cm; Superose 12 Prep Grade, GE Healthcare) that had been equilibrated with 20 mM sodium phosphate (NaPi), 100 mM NaCl, pH 6.0. The sample was eluted at 1 mL/min and 1 mL fractions were collected. Fractions with an absorbance ratio at 410 nm/280nm of greater than 1.2 were pooled and concentrated to ~0.8 mM. The aliquots were flash frozen in liquid N₂, and stored at -80 °C. The concentration of CcP was determined using UV-Vis spectroscopy at $\epsilon_{408\text{nm}} = 98 \text{ mM}^{-1}\text{cm}^{-1}$ and the coordination of the haem group was determined using several absorbance ratios.⁽²²⁰⁾ The standard yields were approximately 70 mg/L culture in rich medium and approximately 10, 70 or 130 mg/L culture in minimal medium for [²H¹³C¹⁵N], [¹³C¹⁵N] or [²H ¹⁵N] labelled, respectively.

The mass of the apo-protein as well as the mass of the haem group were determined using mass spectroscopy. Proteins were desalted with P6 size exclusion cartridges (Biorad, Veenendaal, Netherlands) in a 20 mM ammoniumacetate buffer pH 8. The protein solution was mixed 1:1 with a 0.1% formic acid, 50% acetonitrile solution and infused via the nano-flow source at 3 $\mu\text{l}/\text{min}$ in a Synapt G2Si mass spectrometer (Waters, Milford, MA, USA), mass calibrated with GluFib, operating in the resolution mode with the following tuning parameters: spray voltage 3 kV, temp 80 °C, sample cone 25 V, source offset 25, gas flow cone 50 L/hr, nanoflow 0.49, purge gas 250 L/hr.

EPR SPECTROSCOPY

To study the haem iron coordination, EPR samples containing 400 μM triple labelled [^{15}N , ^{13}C , ^2H] CcP (fresh or aged) were prepared in 20 mM NaPi buffer with 100 mM NaCl, 6% D_2O , pH 6.0. A HF bound sample was also produced by adding 100 mM KF. The X-band cw EPR measurements were performed at low temperature using an Elysya E680 spectrometer (Bruker, Rheinstetten, Germany) with a rectangular cavity and a modulation frequency of 100 kHz. For measurements at 15 K, a helium gas flow cryostat (Oxford Instruments, United Kingdom) with an ITC502 temperature controller (Oxford Instruments, United Kingdom) was used. For the measurements in frozen solution, 3 mm outer diameter quartz sample tubes were used. To obtain a frozen glass, 20% glycerol was added to the samples before freezing them in liquid nitrogen. The EPR spectra were recorded using a modulation amplitude 5.0 G and a microwave power of 19.97 mW. Total accumulation time was approximately 5 min.

NMR SPECTROSCOPY

BACKBONE RESONANCE ASSIGNMENT

All NMR samples contained 20 mM NaPi, 100 mM NaCl, 6% D_2O , pH 6.0. The pH of the samples was adjusted to 6.00 ± 0.05 , with small aliquots of 0.5 M HCl or NaOH. To determine the stability of an NMR sample containing 500 μM double labelled [^{15}N , ^{13}C] CcP, 2D [^{15}N , ^1H] TROSY-HSQC^(224, 237-241) spectra were obtained with 1024 and 110 complex points in the direct and indirect dimensions, respectively, recorded at 283 K, 293 K or 303 K at several time points. To determine the optimal CcP concentration, 2D [^{15}N , ^1H] TROSY-HSQC^(224, 237-241) spectra were obtained with 1024 and 128 complex points in the direct and indirect dimensions, respectively, for 400-800 μM double labelled [^{15}N , ^{13}C] CcP at 293 K. Measurements were performed at ^1H Larmor frequencies of 600 MHz on a Bruker Avance III spectrometer equipped with a TCI-Z-GRAD CryoProbe (Bruker, Karlsruhe, Germany). The data were processed using Topspin 3.1 (Bruker, Karlsruhe, Germany) and spectral assignment and analysis was done using CCPN analysis 2.1.5.⁽²⁴²⁾

CcP appears to be stable at 20 °C for only 4-5 days, so several samples were required for the backbone assignment experiments. A large sample of 400 μM triple labelled [^{15}N , ^{13}C , ^2H] CcP was prepared and then aliquoted into several identical samples. To ensure adequate deuterium back-exchange, the sample was left at room temperature for 24 hrs. A full set of protein amide backbone assignment experiments were recorded and processed at the Biomolecular Magnetic Resonance facility, Goethe University, Frankfurt by Dr. Frank Löhr.

The data were processed using Topspin 3.1 (Bruker, Karlsruhe, Germany) and spectral assignment and analysis was done using CCPN analysis 2.1.5.⁽²⁴²⁾

All spectra were recorded at a sample temperature of 293 K using Bruker Avance spectrometers equipped with $^1\text{H}[^{13}\text{C}/^{15}\text{N}]$ cryoprobes. TROSY-type triple-resonance experiments⁽²⁴³⁾ were carried out as BEST versions^(244, 245) at ^1H Larmor frequencies of 950 MHz for HN(CA)CB, HNCA, HN(CA)CO and HNCO and 600 MHz for HN(CO)CA and HN(CO)CACB (see table 3.1 for number of complex points).

TABLE 3.1 Numbers of complex points used to acquire the spectra used for the amide backbone assignment of CcP.

Experiment	^1H complex points	^{15}N complex points	^{13}C complex points
BEST-TROSY-HSQC	768	384	-
BEST-TROSY-HN(CA)CB	512	128	112
BEST-TROSY-HNCA	512	128	64
BEST-TROSY-HNCO	512	128	64
BEST-TROSY-HN(CA)CO	512	128	64
BEST-TROSY- HN(COCA)CB	384	100	80
BEST-TROSY- HN(CO)CA	384	100	28

The 3D HN(CA)CB, HNCA, HN(CA)CO and HN(COCA)CB experiments were acquired with non-uniform sampling using 25-30% of the full grid in the indirect dimensions. A shaped sample tube was employed at 950 MHz due to the high salt concentration in the sample. 2D BEST-TROSY-HSQC experiments^(244, 245) were acquired before and after each experiment to ensure the sample had not degraded. These spectra were recorded and processed at the Biomolecular Magnetic Resonance facility, Goethe University, Frankfurt. The data were processed using Topspin 3.1 (Bruker, Karlsruhe, Germany) and spectral assignment and analysis was done using CCPN analysis 2.1.5.⁽²⁴²⁾ The protein backbone assignment was deposited as BMRB entry 19884.⁽¹⁴⁰⁾

These spectra contained double peaks, separated by PCS, with similar intensities for many of the residues indicating the presence of two forms (see tables S1 and S2 for chemical shifts and assignments of the two states). Using a titration from pH 5.9 to 8.0, the different change in peak intensities was used to distinguish the two forms. Fresh samples of CcP gave only one set of peaks in BEST-TROSY-HSQC spectra, matching those distinguished as the major form. The $\Delta\chi$ -tensors were determined in Numbat⁽⁹⁶⁾ using the PCS data and the structure of CcP form PDB entry 2GB8,⁽¹⁾ which is based on the crystal structure of the

complex (PDB entry 2PCC).⁽¹⁶⁹⁾ The location of the paramagnetic centre was determined using Numbat⁽⁹⁶⁾ or was fixed at the position of the iron ion. The assignments for iron-free CcP (BMRB 19075)⁽²¹⁶⁾ were used to obtain the PCS and calculate the individual $\Delta\chi$ -tensors for the two forms (see table S3 for PCS values). Using the $\Delta\chi_{ax}$ values for the two $\Delta\chi$ -tensors, the zero-field splitting constant, D , was determined according to equation 3.1 (see list of symbols at beginning of thesis for all symbol descriptions):⁽²⁴⁶⁾

$$\Delta\chi_{ax} = -D \left(\frac{S(S+1)(2S-1)(2S+3)g_e^2\beta^2}{30k^2T^2} \right) \quad 3.1$$

1D NMR

To study the resonances strongly affected by the paramagnetic centre, NMR samples containing 400 μM CcP (fresh or aged) were prepared in 20 mM NaPi buffer with 100 mM NaCl, 6% D_2O , pH 6.0. The pH of the samples was adjusted to 6.00 ± 0.05 , with small aliquots of 0.5 M HCl or NaOH. The ^1H -WEFT NMR experiments⁽²⁴⁷⁾ were obtained with 4096 points, 204800 scans and a spectra width of 198.3688 ppm at 293 K. Measurements were performed at ^1H Larmor frequencies of 600 MHz on a Bruker Avance III spectrometer equipped with a TCI-Z-GRAD CryoProbe (Bruker, Karlsruhe, Germany). The data were processed using Topspin 3.2 (Bruker, Karlsruhe, Germany). Due to the large sweep width necessary to observe all the isotropic shifts, the baseline had to be corrected using the polynomial baseline correction function built into the Topspin software.

To test for the presence of fluoride contamination, ^{19}F NMR was done on several samples, each containing a single stock compound in the concentration used during the preparation of the CcP NMR sample including hemin, Na_2HPO_4 , NaH_2PO_4 , NaCl, NaOH, HCl, D_2O and MQ with a sample of 100 mM KF as a positive control. All samples were left at room temperature for one week prior to the measurements to simulate aging. The ^{19}F NMR experiments were obtained with 262144 points, 16 scans and a sweep width of 46948.3 Hz at 297 K. Measurements were performed at ^{19}F Larmor frequencies of 376.36 MHz on a Bruker DMX400 spectrometer equipped with a 5 mm Sel [$^1\text{H}/^{19}\text{F}$] probe (Bruker, Karlsruhe, Germany). The data were processed using Topspin 1.3 (Bruker, Karlsruhe, Germany).

RESULTS & DISCUSSION

PROTEIN SAMPLE PREPARATION

The triple resonance NMR experiments used for protein backbone assignment require large amounts of isotope-labeled protein. Therefore, optimization of the expression and purification protocol was necessary. The first step was to clone the yeast CCP1 gene (with mutation C128A) into a pET28a(+) vector that utilizes KAN resistance selection which is ideal for slow growing cultures in minimal media.^(248, 249) The plasmid controls expression of CCP1 using a T7lac promoter, which can only be bound by T7 polymerase. The gene for T7 polymerase⁽²⁵⁰⁾ is present in the *E. coli* strain used for expression, BL21 Star (DE3)pLysS. These cells carry the gene for the T7 polymerase⁽²⁵⁰⁾ under control of the inducible lacUV5 promoter.⁽²⁵¹⁾ In order to prevent leaky T7 polymerase expression, the pET28a(+) vector contains a constitutively expressed gene for the lac repressor lacI.^(248, 249) In addition, the BL21 Star (DE3)pLysS cells also constitutively express T7 lysozyme, which binds T7 polymerase and inhibits activity prior to induction with IPTG. Upon induction, the rate of T7 polymerase expression becomes much higher than that of T7 lysozyme, making the amount of T7 lysozyme present insufficient to inhibit T7 polymerase activity.⁽²⁵²⁾ This allows for extremely high, yet tightly controlled levels of expression. Furthermore, these cells contain a mutation in the RNase E gene, which enhances RNA stability and thereby protein expression.^(253, 254)

For the purification protocol, all buffers were changed to NaPi buffers since acetate can bind to and even react with the haem group.⁽¹⁴⁶⁾ The standard yields for this protocol are approximately 70 mg/L culture in rich medium and approximately 10, 70 or 130 mg/L culture in minimal medium for [²H¹³C¹⁵N], [¹³C¹⁵N] or [²H ¹⁵N] labelled CcP, respectively.

Originally, CcP was reported to have very poor solubility, less than 10 μ M in water.⁽²¹⁵⁾ However, the construct produced with this protocol is stable at concentrations as high as 1.4 mM for short periods of time and at concentrations between 0.8-1.0 mM for long term storage. The previously published solubility was measured in distilled water so the use of phosphate buffer containing 100 mM NaCl buffer probably contributes to this enhanced solubility. Furthermore, it should be noted that the N-terminal residues of this CcP construct have been changed to MSKT;⁽¹⁴⁰⁾ it is possible that this contributes to enhanced stability/solubility of the protein as well.

In addition to being sufficiently soluble for the NMR experiments, it was also necessary for CcP to be reasonably thermo-stable. For protein NMR, there is always a trade-off between high acquisition temperatures, at which the peaks are well defined, and low acquisition temperatures, at which the protein sample is stable. In order to determine the optimal temperature for NMR measurements, 2D [^{15}N , ^1H] TROSY-HSQC spectra were recorded at 10 °C, 20 °C or 30 °C at several time points, between which the samples were stored at the same respective temperatures (Figure 3.1).

Predictably, the protein was found to be the most stable at 10 °C (see Figure 3.2A) and least stable at 30 °C (see Figure 3.2C) as can be seen by the large increase in peaks around 8.0 ppm in the ^1H domain overtime for the sample at 30 °C indicating protein unfolding. However, the peaks were noticeably broader at 10 °C compared to 30 °C. Therefore, a compromise of 20 °C was chosen for further NMR experiments as it provided good quality spectra and the samples were stable at this temperature for 4-5 days. In order to run the full set of 3D resonance assignment experiments, requiring approximately two weeks of measurement time, several identical NMR samples were used successively and 2D spectra were acquired before and after each 3D experiment to ensure the sample had not changed.

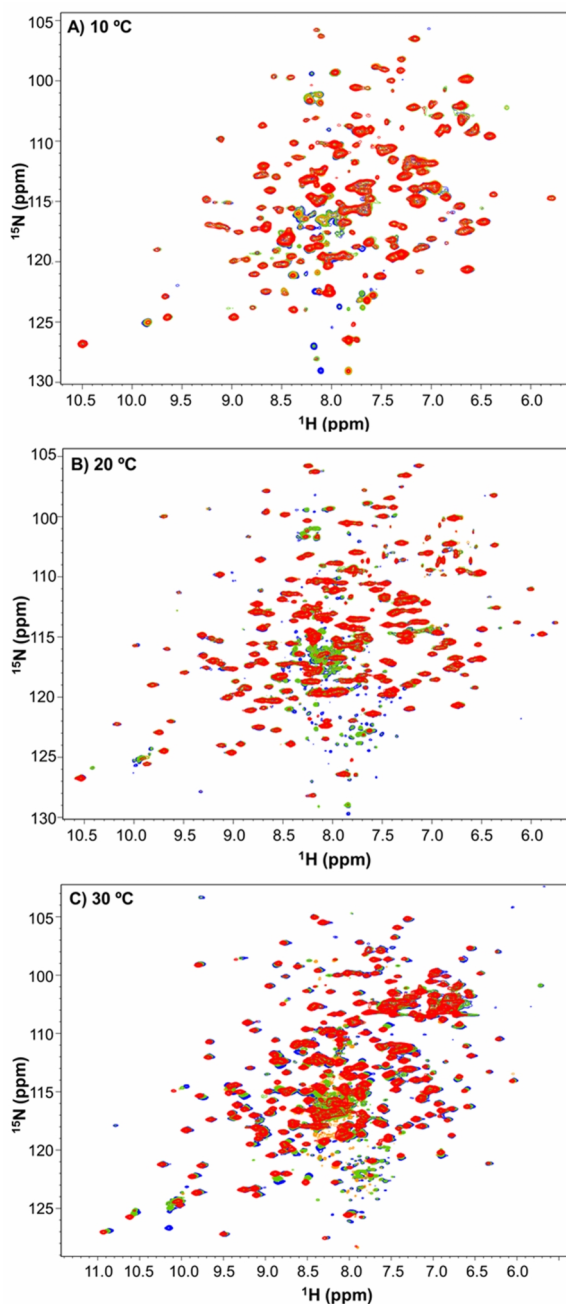


FIGURE 3.1 2D ^{15}N , ^1H TROSY-HSQC spectra of 500 μM double labelled ^{13}C , ^{15}N CcP C128A at 10 °C (**A**) or 20 °C (**B**) for 0 hrs (red), 48 hrs (orange), 96 hrs (green), 168 hrs (blue) or 30 °C (**C**) for 0 hrs (red), 96 hrs (orange), 168 hrs (green), 336 hrs (blue). Experiments were performed in 20 mM NaPi, 100 mM NaCl, pH 6.0.

PROTEIN BACKBONE ASSIGNMENT

A full set of protein amide backbone assignment experiments were recorded for CcP, which allowed for the backbone of CcP to be assigned (BMRB 19884; see Table A1 for chemical shifts in Appendix I).^(140, 255) During the assignment process, an independent assignment of CcP was published with 197 assignments, five of which were used to complement this data set.⁽²¹⁶⁾ Excluding prolines, CcP has 280 assignable residues and, of these, 239 amides (85%) were assigned (Figure 3.2).

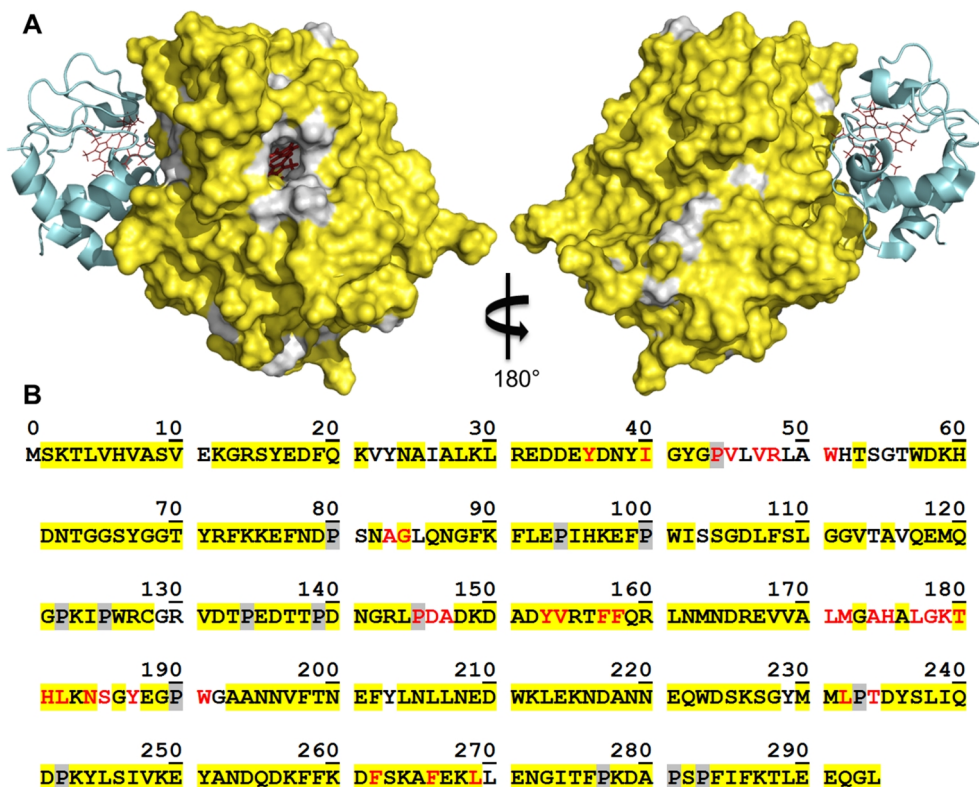


FIGURE 3.2 Extent of the NMR resonance assignments of backbone nuclei of CcP. **A)** Assigned (yellow) and unassigned (grey) residues mapped onto a surface model of CcP (haem group in red sticks) in the stereo-specific complex (PDB-entry 2GB8).⁽¹⁾ Cc is shown in cyan ribbons with the haem group in red lines. **B)** Assigned residues are highlighted in yellow, prolines are highlighted in grey and amides within 5 Å of the haem are shown in red. 85% or 239 of 280 assignable backbone amides (excluding prolines) were assigned. Figure adapted from ref. ⁽¹⁴⁰⁾

Most of the 41 unassigned residues were either buried in the protein, probably experiencing incomplete back-exchange of the deuterons,⁽¹⁴¹⁾ or were within 5 Å of haem iron ion. Furthermore, BEST versions of the assignment NMR experiments were used;^(244, 245) this greatly decreased the required measurement time but also meant that peaks outside of approximately 6.2-10.4 ppm in the ^1H dimension were lost. Nonetheless, a sufficient coverage of the CcP surface was achieved to allow for the mapping of interactions with Cc using pNMR (Chapter 4).⁽¹⁴⁰⁾

HAEM IRON $\Delta\chi$ -TENSOR CALCULATION

In the 2D and 3D protein amide backbone resonance assignment spectra of CcP, many of the peaks were doubled in a pattern which resembled PCS pairs (Figure 3.3).

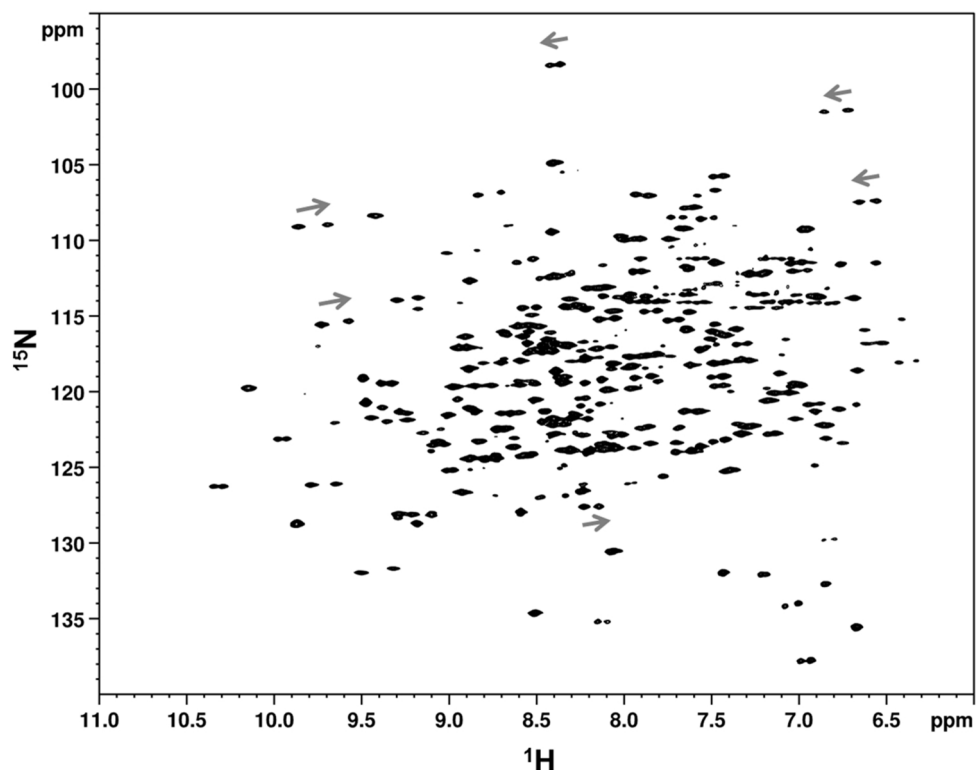


FIGURE 3.3 2D [^{15}N , ^1H] BEST-TROSY-HSQC spectra of 400 μM triple labelled [^2H , ^{13}C , ^{15}N] CcP. The grey arrows indicate a few of the PCS pairs. Experiments were performed at 20 °C in 20 mM NaPi, 100 mM NaCl, pH 6.0.

PCS are generated by the presence of a nearby paramagnetic centre with an anisotropic magnetic susceptibility χ -tensor, such as an iron ion. They are easily recognised because they shift peaks by roughly the same amount in each dimension as can be seen in the 2D-BEST-TROSY-HSQC (Figure 3.3). It must be noted however that in this case, since TROSY NMR experiments were used, the measured shift difference includes the PCS, the RDC and the residual anisotropic chemical shifts (RACS) for the ^{13}C and ^{15}N dimensions.⁽²⁵⁶⁻²⁵⁸⁾ Therefore, although these effects are very small and can be adjusted for during later calculations, only the ^1H PCS values were used for further $\Delta\chi$ -tensor calculations.

Generally, separate diamagnetic and paramagnetic spectra are recorded allowing for determination of the PCS. In this case it was not always clear which peaks belonged to which form as both sets of peaks appeared in the same spectrum. Many of the PCS pairs consisted of a weaker and a stronger peak suggesting a minor and a major state in the sample; although there were several pairs in which the peaks intensities appeared to be the same. In order to determine which peaks belonged to the dominant form, a pH titration was done from 5.9 to 8.0 and the resulting spectra were compared to those used for assignment (Figure 3.4).

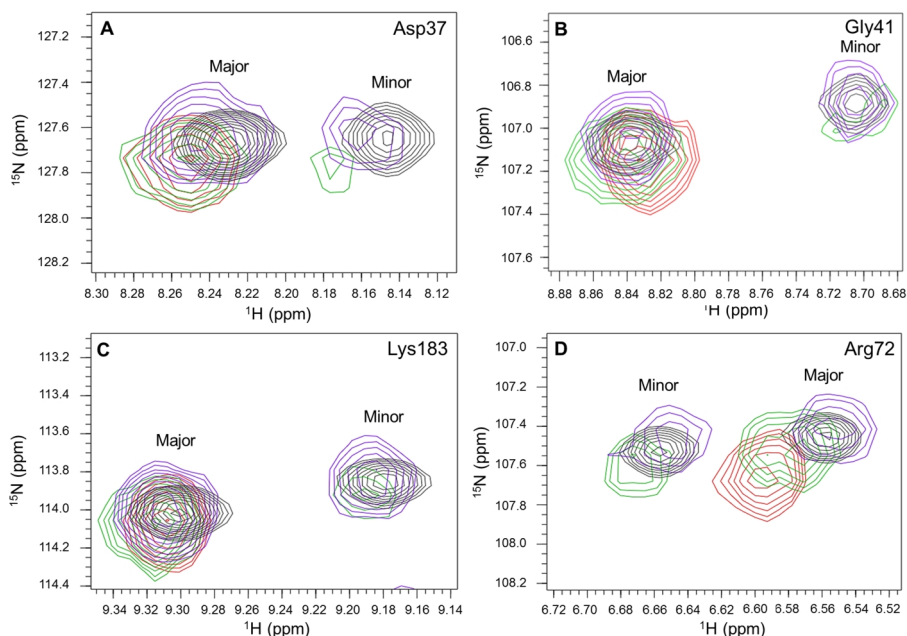


FIGURE 3.4 Close up of peaks for residues Asp 37 (A), Gly 41 (B), Lys 183 (C) and Arg 72 (D) showing the low and high pH forms in 2D [^{15}N , ^1H] TROSY-HSQC spectra of 400 μM triple labelled [^2H , ^{13}C , ^{15}N] CcP at pH 6.0 (used for assignment) (grey), at pH 5.9 (mauve), pH 7.0 (green) and 8.0 (red). Experiments were performed at 20 $^\circ\text{C}$ in 20 mM NaPi, 100 mM NaCl, pH 6.

For solvent exposed amide protons, the H-X exchange rate increases with pH and thus the peak intensity decreases.⁽²³⁷⁾ In this case, both peaks experienced a decrease in intensity as the pH increased; however, the effect was proportionally larger for the weaker peak in the PCS pairs (the major:minor peak intensity ratio increased). This indicated that the proportion of the minor form decreased as the pH increased.

As mentioned above, CcP can exist in three pH dependent states.^(220, 224-229) NMR experiments were conducted at pH 6 and the pH of the samples was checked before and after the NMR experiments to ensure no changes occurred. Nevertheless, the pH dependence of the minor form warranted further investigation to determine if it could be the low pH form, which generally occurs below pH 5.5.^(220, 224, 225) This was done using 1D proton NMR (Figure 3.5).

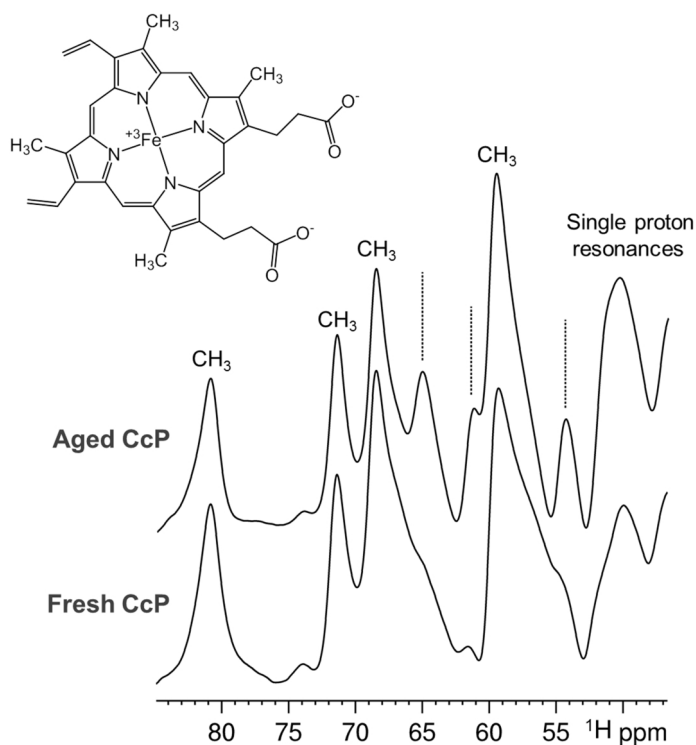


FIGURE 3.5 Detail of the proton NMR spectra of the haem group from fresh and aged 400 μM CcP along with the structure of protoporphyrin IX, the haem prosthetic group. Only a small region of the spectra in which clearly resolvable isotropically shifted resonances appear is shown. Experiments were performed at 20 $^{\circ}\text{C}$ in 20 mM NaPi, 100 mM NaCl, pH 6.

The most characteristic peaks in spectra of haems correspond to the four methyl groups on the protoporphyrin IX ring. For the fresh sample of CcP, these four peaks are clearly visible between 55-85 ppm. In the aged sample of CcP, the peaks are still clearly present, representing the major form of CcP, but additional resonances are visible that we ascribe to the minor form. Satterlee and Erman reported that in the proton NMR spectra of the low pH form of CcP, three of the peaks are also broadened but the peak at 68 ppm is no longer visible.⁽²²⁵⁾ This is clearly not the case here as the 68 ppm peak is shifted slightly in the minor form but remains relatively intense indicating that the minor form observed here is not the low pH form reported elsewhere.

Interestingly, the 1D spectra provide further information about the coordinate state of the haem group. The paramagnetic nature of a HS iron ion results in very large shifts of the peaks to outside the diamagnetic envelope of the ^1H NMR spectrum, especially the characteristic methyl peaks, and these shifts are much larger for hexacoordinate haem groups than pentacoordinate. Conversely, in the spectra of LS haem, the four characteristic methyl peaks are not expected further downfield than about 20 ppm.⁽²⁵⁹⁾ Satterlee and Erman concluded that the native state of CcP in solution is hexacoordinate HS with a water molecule bound in the axial position. The spectra of our fresh CcP matches very well with that published previously for the native enzyme.⁽²²⁵⁾ The small magnitude of the shifts between the methyl peaks of the major and minor form in the spectrum of the aged CcP sample indicate that the minor form is also likely to be hexacoordinate.

Comparison of the amide resonance frequencies of the minor form to those published previously for CcP containing an iron-free protoporphyrin IX haem precursor (PPIX)⁽²¹⁶⁾ or the apo-protein⁽²⁶⁰⁾ further confirmed that the minor form was not likely to be diamagnetic nor did the peaks match those of CN-bound CcP (hexacoordinate LS),⁽¹⁴¹⁾ while the resonance frequencies of the major form matched those published for the RS enzyme.⁽²¹⁶⁾ Therefore, the peak pairs likely represented two HS forms; one for the RS of the enzyme and the other for an unknown minor state. Using the published chemical shifts for PPIX CcP as the diamagnetic position,⁽²¹⁶⁾ PCS could be obtained for both the major and minor forms and individual $\Delta\chi$ -tensors were calculated (Figure 3.6 and Table 3.2).

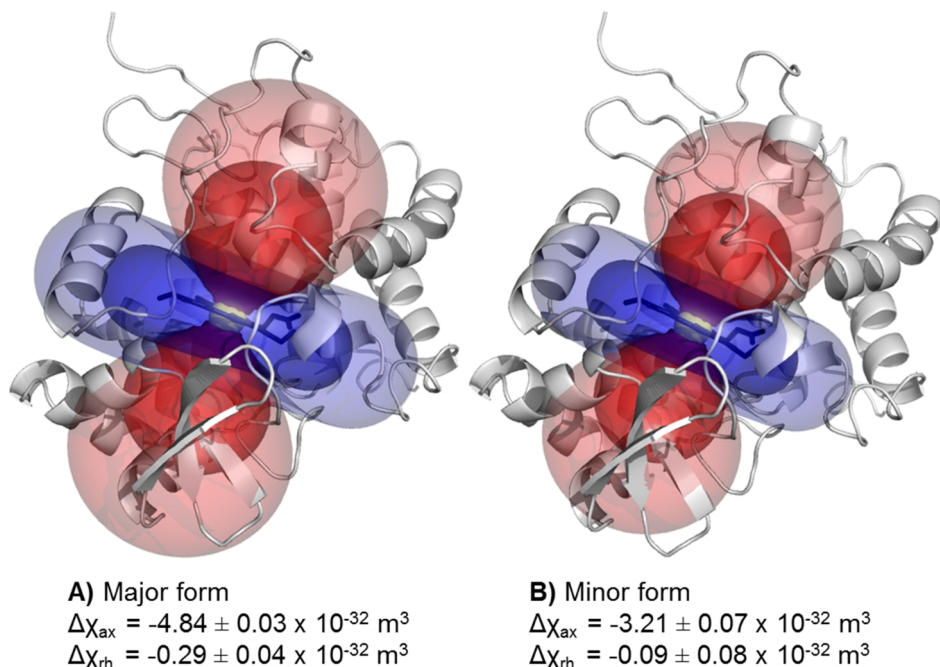


FIGURE 3.6 Isosurface of PCS representing the $\Delta\chi$ -tensor superimposed on the structure of CcP (grey ribbon) and the $\Delta\chi$ -tensor values for the haem iron ion (yellow sphere) of the major (**A**) and minor (**B**) forms of CcP. The blue and red isosurfaces indicate positive and negative PCS of ± 0.700 ppm (dark shade) or ± 0.400 ppm (light shade), respectively. The haem group is shown in black sticks; the paramagnetic centre was fixed on the iron ion. The $\Delta\chi$ -tensor was calculated with Numbat⁽⁹⁶⁾ using PCS obtained from a [¹⁵N, ¹H] BEST-TROSY-HSQC spectrum, the iron-free CcP assignments (BMRB entry 19075)⁽²¹⁶⁾ and the CcP structure (PDB entries 2GB8⁽¹⁾ and 2PCC⁽¹⁶⁹⁾).

TABLE 3.2 $\Delta\chi$ -tensor values for the haem iron ion of the major and minor forms of yeast CcP C128A calculated using Numbat⁽⁹⁶⁾ using PCS obtained from a [¹⁵N, ¹H] BEST-TROSY-HSQC spectrum, the iron-free CcP assignments (BMRB entry 19075)⁽²¹⁶⁾ and the CcP solution structure (PDB entries 2GB8⁽¹⁾ and 2PCC⁽¹⁶⁹⁾) with the paramagnetic centre either fixed on the iron ion coordinates or free.

Major form		
Tensor coordinates	Fe Fixed	Fe Free
$\Delta\chi_{ax}$ (10^{-32} m^3)	-4.84 ± 0.03	-4.83 ± 0.03
$\Delta\chi_{rh}$ (10^{-32} m^3)	-0.29 ± 0.04	-0.34 ± 0.04
x (Å)	-1.68 ± 0.00	-1.55 ± 0.05
y (Å)	33.02 ± 0.00	32.95 ± 0.04
z (Å)	18.64 ± 0.00	18.83 ± 0.04
Minor form		
Tensor coordinates	Fe Fixed	Fe Free
$\Delta\chi_{ax}$ (10^{-32} m^3)	-3.21 ± 0.08	-3.24 ± 0.06
$\Delta\chi_{rh}$ (10^{-32} m^3)	-0.09 ± 0.07	-0.15 ± 0.08
x (Å)	-1.68 ± 0.00	-1.56 ± 0.14
y (Å)	33.02 ± 0.00	33.26 ± 0.00
z (Å)	18.64 ± 0.00	18.86 ± 0.00

Unlike paramagnetic experiments using external tags, in which the position of the paramagnetic centre is unknown, the exact location of the iron ion in the solution structure was known from the PDB file. Therefore, the position of the paramagnetic centre was fixed according to the iron ion coordinates during the calculation of the $\Delta\chi$ -tensors. For the major state, the $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ values were $-4.84 \pm 0.03 \times 10^{-32} \text{ m}^3$ and $-0.29 \pm 0.04 \times 10^{-32} \text{ m}^3$, respectively, which were the same within error of the previously published values for the RS enzyme of $-4.87 \pm 0.05 \times 10^{-32} \text{ m}^3$ and $-0.25 \pm 0.06 \times 10^{-32} \text{ m}^3$, respectively.⁽²¹⁶⁾ The $\Delta\chi$ -tensor components for the minor state were similar in orientation and magnitude to those of the major state, only of smaller magnitude. The $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ values were $-3.21 \pm 0.07 \times 10^{-32} \text{ m}^3$ and $-0.09 \pm 0.08 \times 10^{-32} \text{ m}^3$, respectively. These values matched very well with those expected for HS haem iron $\Delta\chi$ -tensors (both $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ values were negative and the $\Delta\chi$ -tensor was almost completely axial) as compared to what would have been expected for a LS haem iron ion (both $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ values positive and a greater $\Delta\chi_{\text{rh}}$ contribution), suggesting that indeed both forms are HS.⁽²¹⁶⁾ When the position of the paramagnetic centre was not fixed but was rather determined based on the PCS data, its calculated position was very similar to that of the iron ion in the crystal structure (x, y, z coordinates in Table 3.3), indicating that the position is well defined by the data. Also, for both $\Delta\chi$ -tensors, the back-predicted PCS matched very well with the experimental PCS indicating a reliable calculation (Figures 3.7 and 3.8).

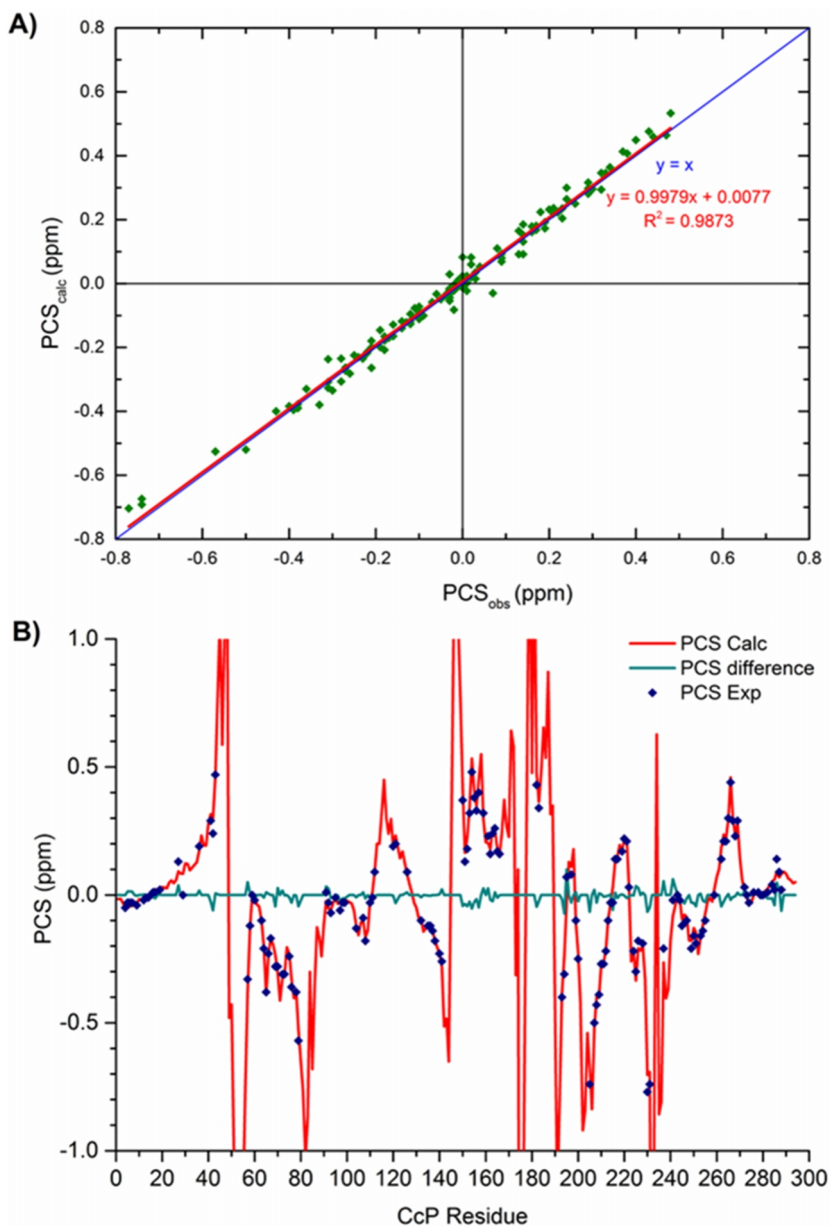


FIGURE 3.7 Comparison of the back-calculated and experimentally observed PCS on CcP amide hydrogen atoms for the $\Delta\chi$ -tensor of haem iron for the major CcP form. **A)** Back-calculated PCS plotted against experimentally observed PCS. The linear best fit is shown in blue and a $y=x$ line is shown in red. **B)** Experimentally observed PCS values plotted on top of the back-calculated PCS values for each CcP residue. The difference between the two is plotted in green. The $\Delta\chi$ -tensor was calculated with Nubat⁽⁹⁾ using PCS obtained from a [^{15}N , ^1H] BEST-TROSY-HSQC spectrum and the CcP solution structure with the iron ion position fixed according to the PDB coordinates (PDB entries 2GB8 and 2PCC).⁽¹⁾

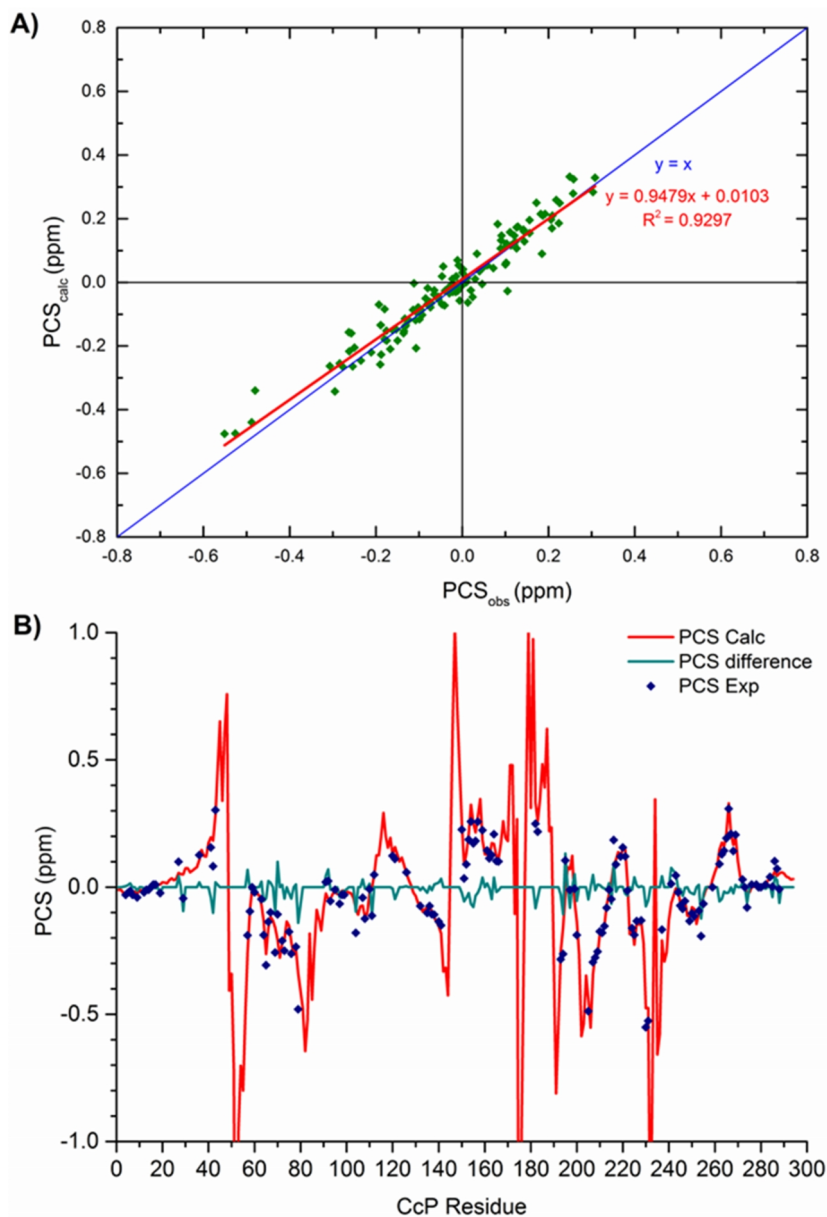


FIGURE 3.8 Comparison of the back-calculated and experimentally observed PCS on CcP amide hydrogen atoms for the $\Delta\chi$ -tensor of the haem iron for the minor CcP form. **A)** Back-calculated PCS plotted against experimentally observed PCS. The linear best fit is shown in blue and a $y=x$ line is shown in red. **B)** Experimentally observed PCS values plotted on top of the back-calculated PCS values for each CcP residue. The difference between the two is plotted in green. The $\Delta\chi$ -tensor was calculated with Numbat⁽⁹⁾ using PCS obtained from a [^{15}N , ^1H] BEST-TROSY-HSQC spectrum and the CcP solution structure with the iron ion position fixed according to the PDB coordinates (PDB entries 2GB8 and 2PCC).⁽¹⁾

When the back-calculated and experimentally observed PCS are plotted against each other for the major form (Figure 3.6A), a very good correlation is obtained ($R=0.99$). A similar result was observed for the minor state (Figure 3.7A) with a slightly weaker but still strong correlation between the experimental and back-calculated data ($R=0.93$). The error in the PCS measurements was very small, on average ± 0.003 ppm (Table S3), which was much smaller than the spread in the data, approximately 0.1-0.2 ppm (Figure 3.6A and 3.7A). This indicates that the measurements are very precise and that most of the deviation is likely due to differences between the haem ligands between the diamagnetic CcP and the paramagnetic forms. These can cause substantial chemical shift perturbations (CSP) in addition to the PCS, especially for residues close to the haem.^(93, 261) Similar results were seen when the back-calculated and experimentally observed PCS were compared for all CcP residues (Figs. 3.6B and 3.7B). Small deviations occurred but they were spread across the sequence and no obvious areas of strong mismatch could be identified.

IDENTIFICATION OF THE MINOR FORM OF CCP

In order to determine if changes in the covalent structure of CcP were the cause of the two forms in the NMR spectra, the mass of the $^2\text{H}^{13}\text{C}^{15}\text{N}$ CcP sample used for the NMR assignment experiments, as well as a fresh sample from the same batch that had not been used for NMR, was measured using mass spectrometry. The theoretical mass for the haem group, without the iron ion, is 616.1773 g/mol, which was almost the same as that found for the both of the CcP haem samples, 616.1906 g/mol. For the protein, the theoretical mass was expected to be close to but slightly greater than that for $^{13}\text{C}^{15}\text{N}$ CcP of 35413.186 g/mol since the protein used was also partially deuterated. For the fresh sample of $^2\text{H}^{13}\text{C}^{15}\text{N}$ CcP, the mass was 37492 g/mol, which was almost identical to that of the sample used for the NMR experiments, 37493 g/mol. Therefore, no evidence was found for covalent modifications of the protein or the haem molecule that could be causing the differences observed between the two samples and thus the differences were most likely caused by two different axial ligands bound to the haem.

Depending on the experimental conditions, the ferric haem iron ion can exist in a pentacoordinate HS, hexacoordinate HS or a hexacoordinate LS state, all of which are paramagnetic.⁽¹⁴⁶⁾ The majority of studies done on CcP have used a protocol in which the last step of purification is crystallization of the enzyme. The crystals are stored until needed at which time they are dissolved in aqueous buffer; this is the “freshly prepared” enzyme.^(139, 142, 146, 220, 262, 263) Freshly prepared CcP is predominantly pentacoordinate HS (equatorially

coordinated to the four pyrrole nitrogens of haem itself and axially coordinated to the N_2 of H175)⁽¹⁴⁴⁾ and it converts to hexacoordinate HS as it ages, in which the haem is coordinated to an axial weak ligand according to ligand field theory, such as water, and has much lower reactivity towards hydrogen peroxide.^(146, 232) The rate at which the sixth ligand binds to the haem is dependent on the duration of storage of the sample prior to use; the sixth ligand can take several hours to bind for a fresh sample or this can occur within the mixing time for a relatively aged sample, suggesting that storage induced structural changes occur around the haem group making the ligand binding more favourable.⁽¹⁴⁶⁾ Which form is present can be determined in several ways but this is most commonly done by measuring an increase in the absorptivity of the Soret band (408 nm) and a shift in the charge transfer band maximum from 645 nm to 620 nm in the UV-Vis spectrum for hexacoordinate compared to pentacoordinate HS haem.⁽²²⁰⁾ The increase in absorptivity of the Soret band ($\epsilon \geq 120 \text{ mM}^{-1}\cdot\text{cm}^{-1}$) hexacoordinate HS haem also leads to higher purity indexes ($A_{408\text{nm}}/A_{280\text{nm}}$); CcP prepartes are considered pure with a ratio ≥ 1.2 while we routinely observed values ≥ 1.4 .⁽¹⁴⁶⁾

The UV-Vis spectra were measured for the sample used for NMR assignment (aged CcP), a fresh sample and a sample containing HF bound CcP, the reason for which will be discussed further below (Figure 3.9 and Table 3.3).

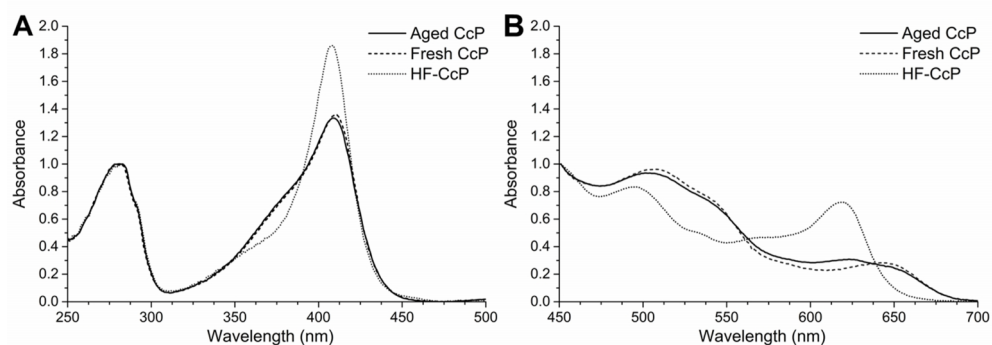


FIGURE 3.9 UV-Vis spectra for CcP for aged (containing both the major and minor forms; black line), fresh (dotted line) and HF bound (grey line) yeast CcP C128A. **A)** Detail of UV-Vis spectra from 250-500 nm normalized based on the absorbance at 280 nm. **B)** Detail of UV-Vis spectra from 450-700 nm normalized based on the absorbance at 450 nm.

TABLE 3.3 UV-Vis absorbance ratios for aged (containing both the major and minor forms), fresh and HF bound yeast CcP C128A compared to the theoretical values⁽²²⁰⁾ for CcP containing either a pentacoordinate or hexacoordinate haem group.

	Absorbance ratio		
	410nm/280nm	410nm/380nm	620nm/645nm
Aged CcP	1.40	1.60	1.62
Fresh CcP	1.34	1.72	0.84
CcP-HF	1.75	2.70	3.00
5-coordinate theoretical	≥ 1.2	1.52	0.74
6-coordinate theoretical	1.2-1.4	2.2-2.5	1.8-3.9

For the aged sample, the 408 nm/380 nm ratio, which measures the increase in absorptivity of the Soret band, was 1.60; this is very close to the value that has been reported for a pentacoordinate HS haem, 1.52.⁽²²⁰⁾ However, the charge transfer band showed a clear shift to 620 nm resulting in a 620 nm/645 nm ratio of 1.62 (0.74 and 1.8-3.9 are characteristic for pentacoordinate and hexacoordinate haem, respectively).⁽²²⁰⁾ While these results are difficult to interpret, the proton NMR spectra (Figure 3.5) strongly suggest that both the major and minor forms of the protein are hexacoordinate HS species.

In HS haem systems, the time-averaged anisotropic component of the $\Delta\chi$ -tensor, which causes the PCS effect, is generated by zero-field splitting, which is the splitting of energy levels in the absence of a magnetic field. This can be a sensitive probe for the haem coordination state.⁽²⁶⁴⁾ Thus, it is possible to ascertain some information about the nature of the axial ligand using the value of the observed zero field splitting constant (D), which can be calculated from the axial component of the $\Delta\chi$ -tensor derived from the PCS analysis (see equation 3.1). The D value for a HS hexacoordinate haem is indeed expected to be around 10 cm^{-1} ⁽²⁶⁵⁾ and as the axial ligand field is decreased, as would be the case for pentacoordinate HS haems, D generally increases. For example, the D values for metmyoglobin hexacoordinated to water range from 9.1 to 9.7 cm^{-1} ⁽²⁶⁶⁻²⁶⁸⁾ while the values for pentacoordinate HS metmyoglobin are 15.8 cm^{-1} (WT) or 13.8 cm^{-1} (His(E7)Val mutant).⁽²⁶⁹⁾

For the major form, D was found to be $10.15 \pm 0.06 \text{ cm}^{-1}$. This was the same as previously published for the RS enzyme indicating a hexacoordinate HS haem (10.2 cm^{-1})⁽²¹⁶⁾ and was also comparable to the values for metmyoglobin hexacoordinated to water obtained from NMR studies (9.7 cm^{-1})⁽²⁶⁶⁾, IR studies (9.5 cm^{-1})⁽²⁶⁷⁾ or through the temperature dependence of the electron spin-lattice relaxation rate (9.3 cm^{-1} for single crystal; 9.1 cm^{-1} for polycrystalline solution).⁽²⁶⁸⁾ The D value for the minor state of CcP was found to be $6.73 \pm 0.15 \text{ cm}^{-1}$, indicating not only that it was unlikely to be pentacoordinate, but also that it was not likely to be bound to water. Similar D values have been found for fluoride bound to other

haem irons: 6.3 cm^{-1} ⁽²⁶⁷⁾ for fluoro-hemoglobin and $5.9\text{-}6.2 \text{ cm}^{-1}$ for fluoro-myoglobin.^(267, 268) Although we did not add fluoride to the sample, the NMR buffers contained sodium chloride. Chloride is also known to bind to the haem iron, especially at low pH,⁽¹⁴⁶⁾ and would be expected to give a slightly higher D value than fluoride based on measurements done on iron bound PPIX.⁽²⁶⁷⁾

To test whether the ligand in the minor form could be chloride or fluoride, EPR spectra of both the fresh and aged samples were obtained (Figure 3.10).

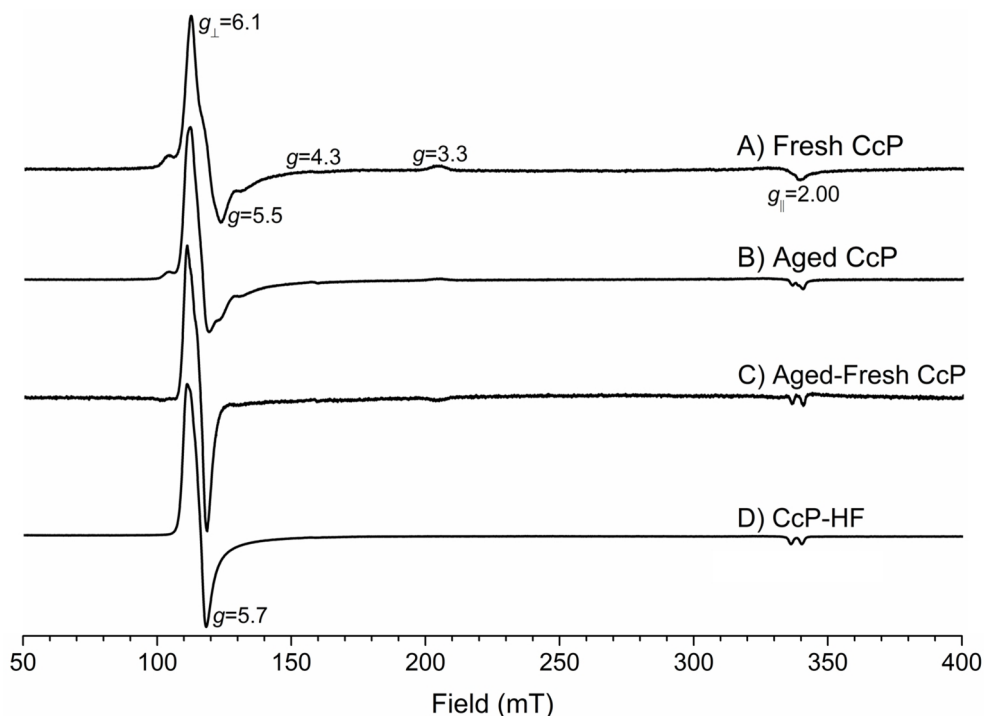


FIGURE 3.10 Comparison of fresh and aged CcP using X-band EPR at 15 K (black lines). **A)** Fresh CcP sample; **B)** aged CcP sample containing a mix of fresh and aged CcP; **C)** Simulated spectra of pure aged CcP obtained by subtracting spectra A from B; **D)** CcP with HF bound to the haem. Microwave frequency: 9.50 GHz; modulation amplitude: 5 G; microwave power: 20 mW.

First, the presence of the characteristic g_{\perp} and g_{\parallel} peaks ($g=6.1$; $g=2$) in the EPR spectra of the fresh and aged samples confirm that both forms of CcP contained a HS spin haem (Figure 3.10 A & B). A LS haem would have had three peaks in the spectrum around $g_x=2.7$, $g_y=2.2$ and $g_z=1.8$ ^(146, 265) and an intermediate spin ($S=3/2$) haem would have had two clear peaks around $g=4.0$.⁽²⁶⁵⁾ The peak at $g=4.3$ represents a small amount of free iron in the samples⁽²⁷⁰⁾ and the peak at $g=3.3$ is the asymmetric zero-crossing resonance.⁽²⁶⁵⁾ The spectrum of the aged sample contains a mix of two states, fresh and aged CcP, so a spectrum of pure aged CcP was generated by subtracting the fresh CcP sample spectrum from that of the aged CcP sample (Figure 3.10 C). The main differences between the fresh and aged CcP spectra can be seen in the peaks at $g=5.5$ and $g=2.0$; in the aged sample, the peak at $g=5.5$ is shifted to $g=5.7$ and the peak at $g=2.0$ is split. The EPR spectrum of a fresh sample with added fluoride (Figure 3.10D) looks remarkably similar to the difference spectrum of the aged CcP sample (Figure 3.10C). The splitting at $g=2.0$ has also been observed in EPR spectra of hemin fluoride and ferrimyoglobin fluoride⁽²⁷¹⁾ and is due to the hyperfine coupling between the electron spin and the nuclear spin of the fluoride, whose main stable isotope ^{19}F has a spin of $I=1/2$. Chloride has a spin of $I=3/2$ and thus the EPR spectrum of hemin chloride should have a splitting into four rather than two.⁽²⁷¹⁾ Therefore, it is indeed likely that HF is the axial ligand in the minor form, which matches well with the observed D values, discussed above. Furthermore, the UV-Vis ratios for the aged sample fall between those of the fresh and HF bound samples, as would be expected for a mixture of both states. This also matches well with the observation that the proportion of the minor form decreased as the pH increased since HF is the dominant binding species at lower pH.⁽²³¹⁾

To confirm that HF was indeed the axial ligand in the minor form, a fresh sample of CcP was converted to HF bound CcP by the addition of 100 mM KF. The NMR spectra of fresh CcP and HF bound CcP were compared to the assignment spectra that contained both the minor and major forms (Figure 3.11 for full spectra and Figure 3.12 for selected peaks). As seen previously, only the major form was present in the fresh CcP sample. The resonances of HF bound CcP matched those of the minor form of CcP confirming that indeed the minor state was caused by ligation of HF to the haem iron. Using the backbone resonance assignment of the major form, 227 resonances of HF bound CcP were successfully assigned (see Table S2 for chemical shifts). This assignment is available as BMRB entry 26633.

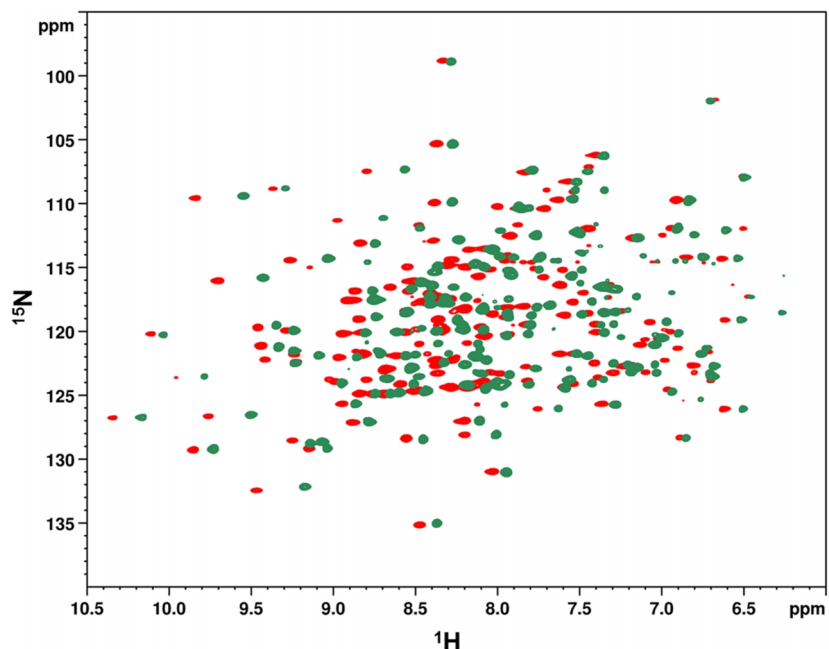


FIGURE 3.11 2D [^{15}N , ^1H] BEST-TROSY-HSQC spectra of fresh (red) or HF bound (green) CcP. Experiments were performed on $400\ \mu\text{M}$ $^2\text{H}^{15}\text{N}$ labeled CcP at $20\ ^\circ\text{C}$ in $20\ \text{mM}$ NaPi, $100\ \text{mM}$ NaCl, pH 6.0 (plus $100\ \text{mM}$ KF for HF bound sample).

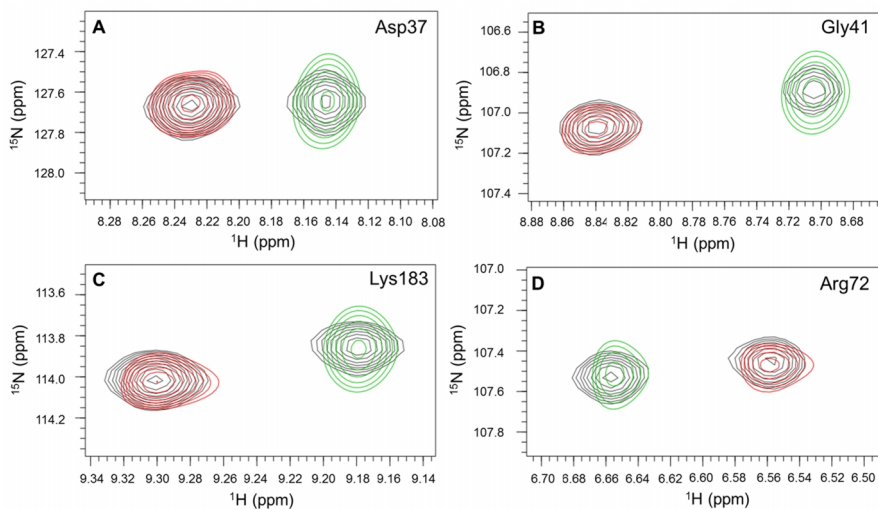


FIGURE 3.12 Close up of peaks for residues Asp 37 (**A**), Gly 41 (**B**), Lys 183 (**C**) and Arg 72 (**D**) showing the 2D [^{15}N , ^1H] TROSY-HSQC spectra of fresh CcP (red) and HF bound CcP (green) overlaid on the spectra used for amide backbone assignment (grey), which contains two forms of CcP; the major form resonances correspond to fresh CcP and the minor form resonances correspond to HF bound CcP. Experiments were performed at $20\ ^\circ\text{C}$ with $400\ \mu\text{M}$ CcP in $20\ \text{mM}$ NaPi, $100\ \text{mM}$ NaCl, pH 6 (plus $100\ \text{mM}$ KF for the HF bound sample).

Although it is clear that the minor state results from HF coordination to the haem group, the origin of this molecule ion is unclear as no fluoride was intentionally added to the samples. Individual samples of the stock compounds used during sample preparation were tested for the presence of fluoride using ^{19}F NMR but no contamination was found. However, the amount of fluoride required to cause the observed effects would be difficult to observe; assuming a maximum conversion of 50% from the major to the minor state, that would require only $0.1\ \mu\text{M}$ ($2\ \mu\text{g}$) of fluoride in the sample. Hydrogen fluoride binds to CcP at pH 5.5 with a rate constant of $1.1 \times 10^8\ \text{M}^{-1}\text{s}^{-1}$ ⁽²³¹⁾ indicating that binding is very fast and cannot be the reason that the fluoride bound state appeared over the course of several days. We conclude that fluoride slowly appeared in the sample, perhaps being released from another compound or solid material, but bound rapidly to CcP once in solution. ⁽²⁷²⁾

The most interesting aspect of this transition was that it appeared to be an artifact of sample aging. The minor state was observed during the long 3D experiments required for assignment and it also began to appear after several days of incubation at $20\ ^\circ\text{C}$ or $30\ ^\circ\text{C}$ (Figure 3.2). It has not since been observed in the NMR spectra of fresh or previously frozen samples, so storage at $-80\ ^\circ\text{C}$ does not seem to contribute to this aging process. This underscores the importance of using fresh samples for every experiment in order to ensure validity of the results.

Here we have described novel expression and purification methods for CcP that produces very high yields of both natural abundance and isotope enriched protein. This resulted in a successful backbone amide resonance assignment of the protein (Chapter 3) ⁽¹⁴⁰⁾ and led to some surprising revelations about the enzyme. Despite most papers describing the haem in the RS of CcP as pentacoordinate HS, ^(139, 142, 146, 220, 262, 263) this is the third NMR study ^(216, 225) showing that fresh CcP in solution is actually hexacoordinate HS with water in the axial ligand position. A second minor state of the protein was also observed in which HF had replaced the water molecule. This replacement appears to happen very slowly during sample aging and has not been seen in any fresh samples; therefore we tentatively ascribe it to contamination from additive release or from the containers used for storage.

CHAPTER 4

The cytochrome *c* peroxidase and cytochrome *c*
encounter complex: The other side of the story

Based on the research article: **Schilder J.**, Löhr F., Schwalbe H. and Ubbink M. (2013). *The cytochrome c peroxidase and cytochrome c encounter complex: the other side of the story*. *Febs Letters* **588**(10): 1873-1878.

ABSTRACT

The encounter state is important for efficient protein complex formation. It consists of an ensemble of orientations of the two proteins in the complex. Experimental description of such ensembles inherently suffers from an insufficient amount of data. The Cc-CcP encounter complex has been described using a set of PRE data obtained with NMR spectroscopy on nuclei in Cc⁽¹⁾ as well as theoretical simulations.⁽⁸⁾ Here, we have tested this model by measuring paramagnetic effects on the larger partner CcP, thus providing a complementary view of the encounter complex. Our data fit the previously published data although also several novel interactions sites were identified. This work demonstrates the need of obtaining as many independent data sets as possible to achieve the most accurate description of an encounter complex.

INTRODUCTION

THE ENCOUNTER COMPLEX AND THE INVERSE PROBLEM

Protein-protein complex formation requires an intermediary complex to form before the final, stereo-specific state is reached. The formation of this encounter complex is driven by long-range charge-charge and hydrophobic interactions, resulting in a weakly associated complex in which the protein partners are free to rotate and reorient themselves. From there, the number of short-range interactions (van der Waals, hydrogen-bonding, hydrophobic interactions and salt bridges) between the pair is increased to form the stereo-specific state.⁽⁵⁷⁾

The transient and highly dynamic nature of the encounter complex makes it difficult to observe and visualize. Because the encounter complex is comprised of a large number of transient, low energy and weakly interacting conformations, it is essentially invisible to many structural biology techniques. PNMR provides a unique opportunity to study these highly dynamic complexes as the observed effects, from PRE in particular, are extremely sensitive for those lowly populated states in which the nucleus is closer to a paramagnetic centre than in the other state(s).⁽⁵⁹⁾

The main drawback is that the PRE, like many other NMR observables, is an average over all the conformations present in the sample. This makes visualization of the complex an *ill-posed inverse problem*,^(273, 274) in which many ensembles of solutions can be found to match the observed data.^(3, 8, 10, 11, 13, 14, 25, 34-36) In fact, the only result that can be determined conclusively is where the interaction does not occur. If a paramagnetic centre does not cause PRE on the partner, it can be concluded that the surface region around that centre is not sampled by the partner for a significant fraction of the lifetime of the complex. Therefore, by using paramagnetic probes at several locations on the protein's surface, an exclusion map can be generated.^(1, 3, 8, 28, 34, 35) The more restraints incorporated into the modelling calculations, the more refined the ensemble of structures becomes and the closer it will be to the true ensemble in the sample.^(32, 33, 275-277)

THE CC-CCP COMPLEX

Encounter complexes are highly populated in complexes that represent a compromise between specific binding and high-turnover. Therefore, electron transport complexes are ideal candidates for studying the encounter complex as they require binding specific enough to allow for ET but weak and transient enough to accommodate very high

turn-over rates.⁽⁷⁾ The ET complex between Cc and CcP is a well characterized system for studying the encounter complex. It spends approximately 30% of the time in the encounter complex,^(1, 8) which can be shifted to as low as 10% or as high as 80% with point mutations near the binding interface.⁽⁹⁾

The solution structure of the Cc-CcP encounter complex was determined in 2006 by Volkov *et al.* using PRE effects generated in the HSQC spectra of Cc by MTSL attached at five locations on the surface of CcP.⁽¹⁾ Although both of these proteins contain a paramagnetic haem group, the effects produced by the haems are not suitable for studying the complex. Therefore, MTSL was used to generate PRE effects, which provided restraints for docking of the proteins. The study demonstrated that the complex spends approximately 70% of the time in the stereo-specific state found in the crystal structure⁽¹⁶⁹⁾ and 30% in other orientations representing the encounter complex. The model of the latter was later refined by Bashir *et al.* in 2010 by expanding the initial data to include PRE restraints from MTSL attached at ten sites on CcP. Back-calculated data from a theoretical encounter complex, generated using an electrostatics based MC simulation, was compared to the experimental PRE data. The additional data obtained allowed for the complete mapping of the conformational space sampled; Cc was found to sample ~15% of the CcP surface during complex formation,⁽⁸⁾ in line with the results from earlier theoretical studies.^(38, 278)

The goal of the present study was to view the Cc-CcP encounter complex from “the other side” and validate the previously determined ensemble. The assignment of the backbone amides of CcP described in Chapter 3 allowed us to observe both CSP and PRE in the NMR spectra of CcP that were generated in the presence of spin-labelled Cc. We observe many effects similar to those previously reported for the complex as well as several novel interactions. These results show the importance of extending the available set of restraints as far as possible to increase the accuracy of an encounter complex description.

MATERIALS & METHODS

EXPRESSION AND PURIFICATION OF CCP

A pET28aCcP plasmid containing the *S. cerevisiae* CcP1 gene with mutation C128A was used to express and purify [¹⁵N²H] labelled CcP as described in Chapter 3. The resulting protein was ~80% deuterated as estimated from ¹D NMR. The concentration of CcP was determined using UV-Vis spectroscopy at $\epsilon_{408\text{nm}} = 98 \text{ mM}^{-1}\text{cm}^{-1}$ and the coordination of the

haem group was determined using several absorbance ratios.⁽²²⁰⁾ The standard yield was approximately 130 mg/L in minimal media.

PROTEIN EXPRESSION AND PURIFICATION OF CC

A pUC19 based plasmid containing the *S. cerevisiae* iso-1-cytochrome *c* gene was used to express and purify Cc as described previously.^(124, 125) The wild type (WT) protein and mutants V28C⁽¹⁴⁾ and A81C⁽²¹⁰⁾ were used. The concentration of Cc was determined using UV-Vis spectroscopy and $\epsilon_{410\text{nm}} = 106.1 \text{ mM}^{-1}\text{cm}^{-1}$.⁽¹²⁵⁾ The standard yield was approximately 20 mg/L in rich media.

SPIN-LABELLING

Samples of V28C and A81C were labelled with either MTS or MTSL [(Toronto Research Chemicals, North York, ON, Canada) as described previously.⁽¹⁾ Purified protein was incubated with 5 mM DTT in 20 mM NaPi, 100 mM NaCl, pH 6.0 for 2 hours at room temperature to reduce intermolecular disulfide bonds. The DTT was then removed by passing the Cc solution through a PD-10 column (GE Healthcare, Hoevelaken, Netherlands). The protein was eluted from the column directly into buffer to which a 10-fold excess of label had just been added. The buffer was incubated at 4 °C with argon bubbled through it for 2 hours prior to the addition of the label and the protein. The solution was stirred gently at 4 °C for 1-2 hours with argon bubbling through it. The sample was then concentrated to 4 mL using Centriprep centrifugal filter devices (Millipore, Amsterdam Zuidoost, Netherlands) and stored at 4 °C overnight. The sample was purified with a gel filtration column (16 cm x 60 cm; Superose 12 Prep Grade, GE Healthcare, Hoevelaken, Netherlands) that had been equilibrated with 20 mM NaPi, 100 mM NaCl, pH 6.0. The sample was eluted at 1 mL/min and 1 mL fractions were collected. The labelling efficiency was determined by mass spectroscopy to be essentially 100%.

NMR SPECTROSCOPY

TITRATION EXPERIMENTS

To obtain binding constants, 1.7-2.5 mM stocks of WT, MTS-V28C or MTS-A81C Cc were titrated into 400 μM $^{15}\text{N}^2\text{H}$ -CcP ($\sim 80\%$ deuterated) in 20 mM NaPi, 100 mM NaCl, 6% D_2O , pH 6.0. 2D BEST-TROSY-HSQC experiments⁽²⁷⁹⁾ were recorded on a Bruker AVIII HD spectrometer equipped with a $^1\text{H}(^{13}\text{C}/^{15}\text{N})$ TCI-cryoprobe operating at a Larmor frequency of 850 MHz at 293 K with 1024 and 100 complex points in the ^1H and ^{15}N dimensions respectively. Spectra were recorded at intervals of 0.2:1 Cc:CcP until a final ratio of Cc:CcP of 2.0:1 was reached. All data were processed using Topspin 3.2 (Bruker, Karlsruhe, Germany) and analysis was done using CCPN analysis 2.1.5.

The average CSP ($\Delta\delta_{\text{avg}}$) were derived as described previously.⁽²⁸⁰⁾ With the derived binding constants, it was calculated that 98% of WT, 99% MTS-V28C Cc or 98% of MTS-A81C Cc was bound to CcP, in the sample with a 2:1 ratio of Cc:CcP. Therefore, in order to obtain $\Delta\delta_{\text{avg}}$ extrapolated to the 100% bound form, the respective $\Delta\delta_{\text{avg}}$ values were divided by 0.98 or 0.99. The chemical shift titration curves were analyzed with a two-parameter, non-linear least squares fit using a one-site binding model as described previously.⁽²⁸¹⁾ The fitting was done using OriginPro 8.5 (OriginLab, Northampton, USA).

PRE EXPERIMENTS

NMR samples contained 400 μM double labelled $^{15}\text{N}^2\text{H}$ -CcP in 20 mM NaPi, 100 mM NaCl, 6% D_2O , pH 6.0 with either 290 μM MTS(L)-V28C Cc or 120 μM MTS(L)-A81C Cc. 2D BEST-TROSY-HSQC experiments were recorded and processed as described for titration experiments. The intensity ratio ($I_{\text{para}}/I_{\text{dia}}$) was determined for all observed amide proton resonances in the spectra of CcP with MTS-Cc (diamagnetic) or MTSL-Cc (paramagnetic) samples (Figure 4.1).

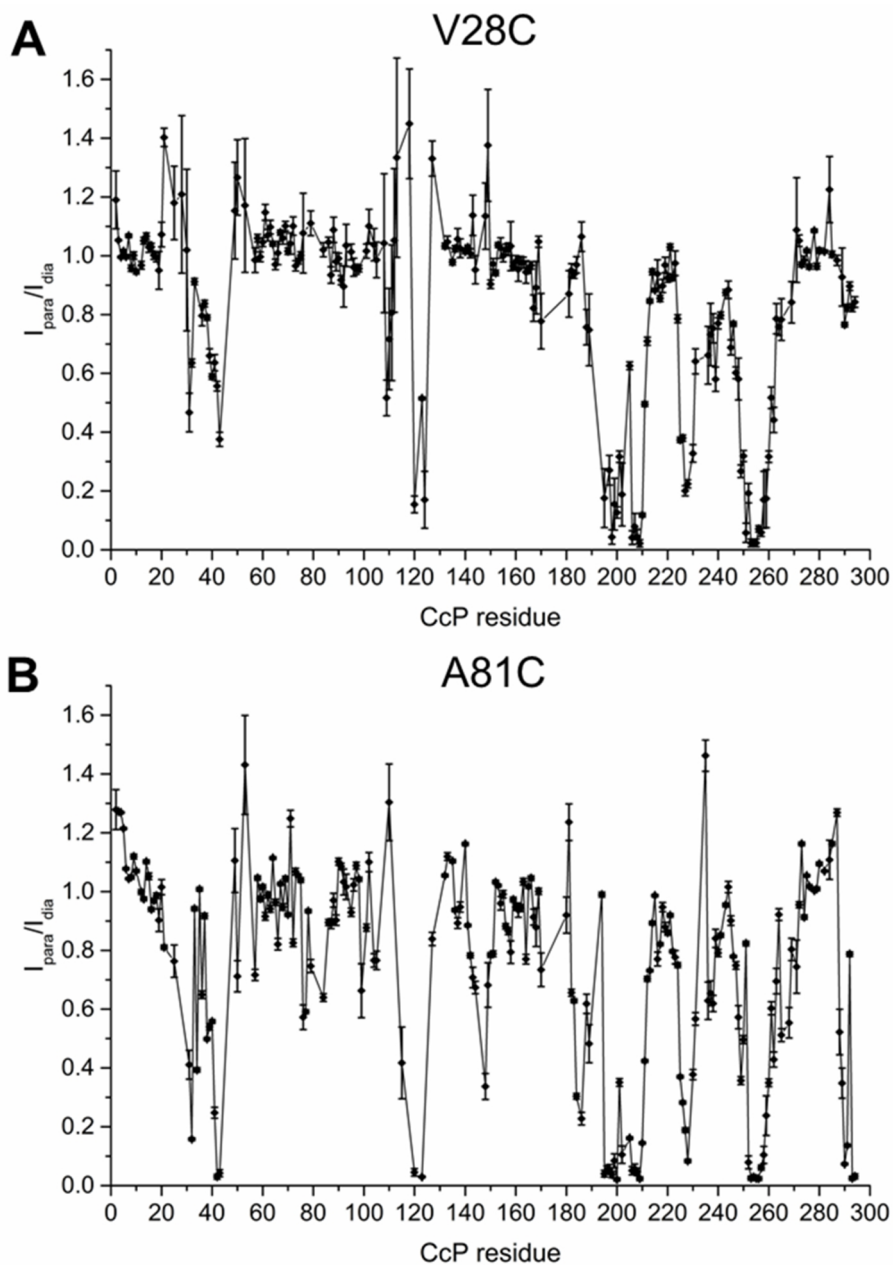


FIGURE 4.1 $I_{\text{para}}/I_{\text{dia}}$ values (circles) for CcP C128A amide protons in the presence of MTS(L) labelled Cc. **A)** $I_{\text{para}}/I_{\text{dia}}$ values in presence of MTSL-V28C Cc with 72% CcP bound. **B)** $I_{\text{para}}/I_{\text{dia}}$ values in presence of MTSL-A81C Cc with 81% CcP bound. The standard deviation of the noise level was determined for each spectrum and was then used in the calculation of the propagated error (bars) for each $I_{\text{para}}/I_{\text{dia}}$ value.

For the A81C data, the I_{para}/I_{dia} values were normalized as done previously.⁽⁸⁾ This was done by estimating the average I_{para}/I_{dia} for unaffected residues (approximately 0.7) and then using the data for residues within approximately 0.2 of that value ($I_{para}/I_{dia} \geq 0.5$). The upper and low 10% of the values were discarded and the average of those remaining was used. The normalization was done by dividing all I_{para}/I_{dia} values by this average.

The $R_{2,para}$ was calculated as described previously.^(8,77) For amides that gave an I_{para}/I_{dia} but for which the line width of the diamagnetic peak could be not obtained, the average value of all the calculated $R_{2,dia}$ values was used with a large error margin. For the amide peaks that disappear in the paramagnetic spectrum, an upper limit for I_{para} was set to two standard deviations of the noise level of the spectrum. The calculated $R_{2,para}$ values were then converted into distances using equation 4.1 (see list of symbols at beginning of thesis for all symbol descriptions):

$$r_{IM} = \sqrt[6]{\frac{f_{bound}}{R_{2,para}} \frac{\gamma_H^2 g^2 \beta^2 \mu_0^2 (S+1) S}{240\pi^2} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right)} \quad 4.1$$

where r_{IM} is the distance between the unpaired electron of the MTSL and a given amide proton of CcP, f_{bound} is the fraction of CcP bound to Cc (0.73 for V28C and 0.81 for A81C) and τ_c was estimated to be 16 ns.⁽¹⁾ The calculated distances were divided into three classes: strongly affected residues for which the peaks had been completely broadened out in the paramagnetic spectrum and only an upper limit could be calculated, affected residues for which the peaks were visible in the paramagnetic spectrum (error margins were set to at least ± 3 Å to account for experimental error) and residues that were too far away from the SL to experience significant PRE, so only a lower limit could be calculated. These distances were then compared to back-calculated distances for a stereo-specific , encounter or 30% encounter/70% stereo-specific complex.⁽⁸⁾

The MTSL was modelled at the Cc mutation sites rotated into three rotamers (A, B and C) for V28C or four rotamers (A, B, C and D) for A81C with widely varying orientations of the MTSL ring (Figure 4.2). Two rotamers (A and B) for V28C and three rotamers (A, B and C) for A81C did not lead to steric clashes with CcP in the stereo-specific complex (PDB entry 2PCC)⁽¹⁶⁹⁾ and were used to calculate the distance between the amide protons of CcP and the oxygen of MTSL for the stereo-specific complex.

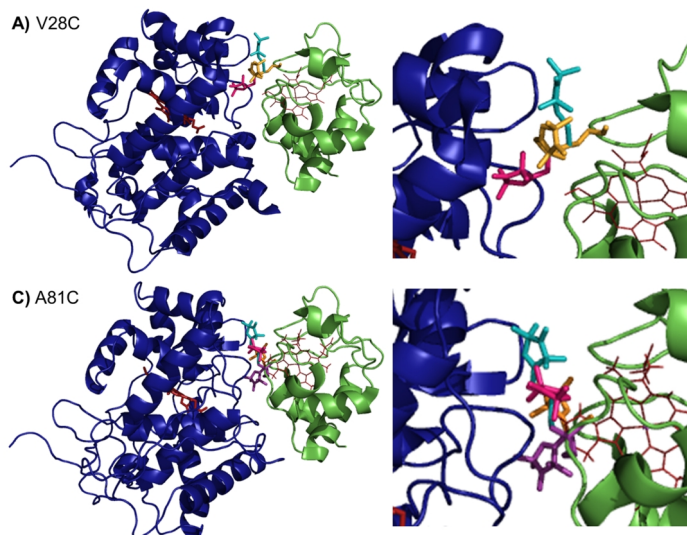


FIGURE 4.2 Orientations of MTSL attached to **(A)** V28C and **(B)** A81C Cc, used to determine back-calculated distances, in complex with CcP (PDB-entry 2GB8).⁽¹⁾ Cc is shown in green ribbons with the haem group in red lines and MTSL is shown in teal (rotamer A), orange (rotamer B), pink (rotamer C) or purple (rotamer D; A81C only). CcP is shown in blue ribbons with the haem group in red sticks.

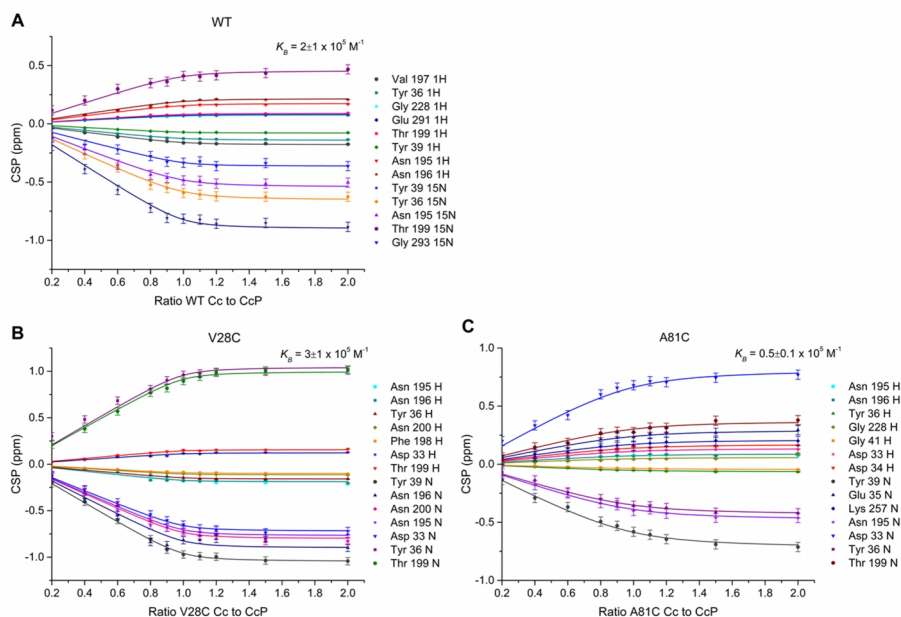


FIGURE 4.3 Chemical shift perturbations for select CcP residues in the ^1H or ^{15}N dimension during titration with WT Cc **(A)**, MTS-V28C Cc **(B)** or MTS-A81C Cc **(C)**. The curves were fitted globally to a 1:1 binding model (see Materials and Methods) and the solid lines show the best fit when using a shared K_B value. These experiments were done in 20 mM NaPi, 100 mM NaCl (pH 6) at 293 K.

All rotamers were used to calculate the distances between the MTSL oxygen and the backbone amide hydrogens in CcP for each of the 1701 orientations of Cc observed in the MC ensemble used to describe the encounter complex previously.⁽⁸⁾ Orientations in which the oxygen approached any of the CcP C_α atoms to less than 5 Å were excluded, because they were taken to represent a penetration of CcP by the SL. For V28C rotamers A, B and C this resulted in 16, 2, and 204 exclusions, respectively. For A81C rotamers A, B, C and D this resulted in 24, 93, 41 and 109 exclusions, respectively. To model the 30% encounter/70% stereo-specific complex, the back-calculated distances from the stereo-specific complexes were combined with the back-calculated distances from the encounter complexes to generate six models for V28C or twelve models for A81C. The r^6 average of these models with an error of two standard deviations was compared to the experimental data. Molecular manipulations and distance measurements were done using XPLOR-NIH.^(282, 283)

RESULTS & DISCUSSION

CHEMICAL SHIFT PERTURBATIONS

The mutations V28C and A81C on Cc were chosen because they are located close to the binding interface of the complex with CcP and MTSL could be modelled into those locations in the crystal structure without resulting in steric clashes. To ensure that attachment of the tag in this location did not significantly disrupt complex formation, WT, V28C-MTS and A81C-MTS Cc were titrated into ¹⁵N²H-CcP and CSP were monitored. Numerous resonances shifted in the spectrum, indicating a fast-exchange binding process. The K_B values were determined for the complexes by fitting the CSP curves to a 1:1 binding model (Figure 4.3). The K_B determined for the complex with WT Cc is $K_B = 2 \pm 1 \times 10^5 \text{ M}^{-1}$, which is the same within error as previously reported.^(60, 183) The binding constant for the complex with MTS-V28C was found to be the same within error, $K_B = 3 \pm 1 \times 10^5 \text{ M}^{-1}$. For MTS-A81C, the binding was found to be slightly weaker with $K_B = 0.5 \pm 0.1 \times 10^5 \text{ M}^{-1}$. As mentioned above, A81C is located to binding interface between Cc and CcP, so it is plausible that the presence of the SL interferes with complex formation resulting in a decreased K_B value. It has been shown that lowering the binding constant of this complex can increase the fraction in the encounter state. In particular, the mutation R13K on Cc ($K_B = 0.54 \pm 0.03 \times 10^5 \text{ M}^{-1}$), which reduces the interactions in the specific state and has a very similar binding constant to MTS-A81C, shifted the equilibrium to 50% encounter/50% stereo-specific complex.⁽⁹⁾ As A81C is in the interface of the specific complex, a similar shift towards the encounter complex may occur here as well.

These K_B values were then used to extrapolate average amide shifts, $\Delta\delta_{\text{avg}}$, for 100% bound CcP (Figure 4.4). For the WT complex, the overall CSP pattern was similar to that described previously.⁽²¹⁷⁾ The CSPs for WT and MTS-V28C or MTS-A81C in this study were also identical within the error margins (± 0.016 ppm) except for a few peaks showing slightly larger differences (Tables 4.1 and 4.2). The $\Delta\delta_{\text{avg}}$ values were also used to create a CSP map (Figure 4.5).

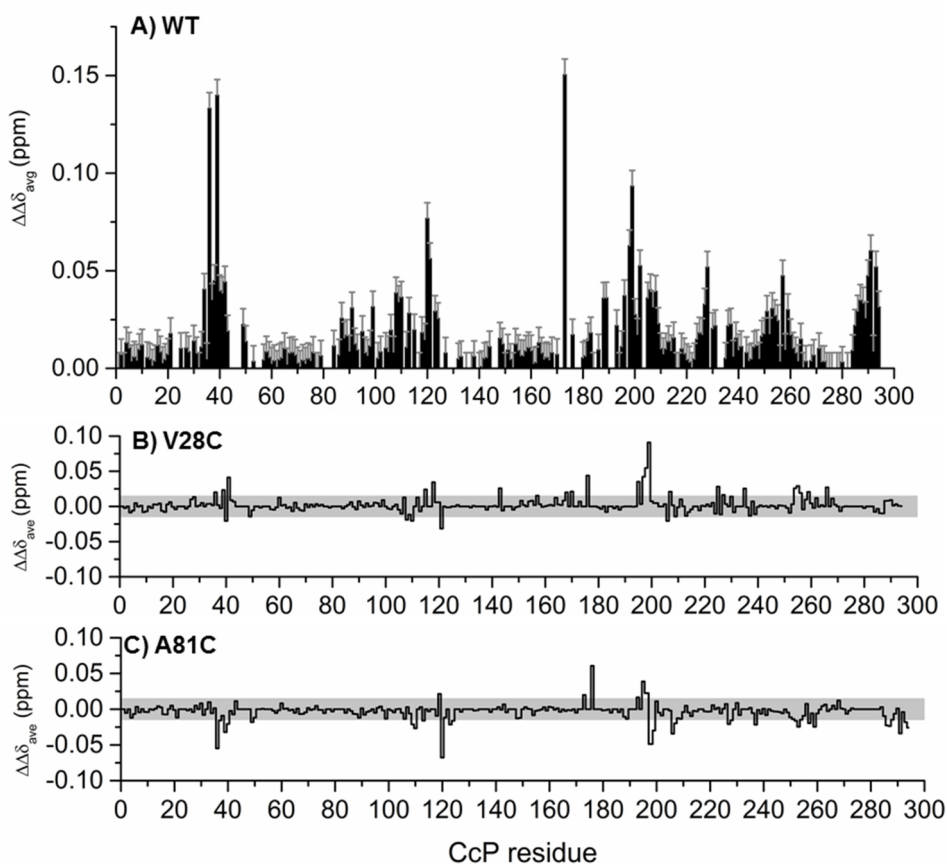


FIGURE 4.4 Chemical shift perturbations for CcP residues upon binding to Cc. **A)** The average amide shift ($\Delta\delta_{\text{avg}}$) for CcP amide groups in the presence of WT Cc in black with grey error bars plotted against CcP residue number. CSP were extrapolated to 100% bound CcP. **B,C)** The difference between $\Delta\delta_{\text{avg}}$ for MTS-V28C and WT (**B**) and for MTS-A81C and WT (**C**) with the error margin in grey for each CcP residue. The experiments were conducted in 20 mM NaPi and 100 mM NaCl (pH 6.0) at 293 K.

TABLE 4.1 Chemical shift perturbation ($\Delta\delta_{\text{avg}}$) for residues with a > 0.016 ppm difference between the complex of WT or MTS-V28C Cc bound to $^{15}\text{N}^2\text{H}$ -CcP C128A. CSP were extrapolated to 100% bound CcP.

CcP residue	WT $\Delta\delta_{\text{avg}}$ (ppm)	MTS-V28C $\Delta\delta_{\text{avg}}$ (ppm)	$\Delta\Delta\delta_{\text{avg}}$ MTS-V28C and WT (ppm)
36	0.133	0.153	0.020
39	0.140	0.163	0.023
40	0.040	0.019	-0.031
41	0.039	0.080	0.041
108	0.039	0.020	-0.019
110	0.036	0.016	-0.020
115	0.020	0.044	0.024
118	0.018	0.053	0.035
121	0.056	0.025	-0.031
143	0.006	0.032	0.026
168	0.008	0.028	0.020
170	0.007	0.028	0.021
176	0.017	0.061	0.044
195	0.011	0.046	0.035
197	0.135	0.177	0.042
198	0.063	0.118	0.055
199	0.093	0.184	0.091
206	0.040	0.020	-0.020
207	0.032	0.053	0.022
225	0.018	0.046	0.028
235	0.005	0.031	0.024
254	0.027	0.053	0.026
255	0.025	0.054	0.029
256	0.004	0.023	0.019
259	0.030	0.051	0.021
266	0.004	0.031	0.027

TABLE 4.2 Chemical shift perturbation ($\Delta\delta_{\text{avg}}$) for residues with a > 0.016 ppm difference between the complex of WT or MTS-A81C Cc bound to $^{15}\text{N}^2\text{H}$ -CcP C128A. CSP were extrapolated to 100% bound CcP.

CcP residue	WT $\Delta\delta_{\text{avg}}$ (ppm)	MTS-A81C $\Delta\delta_{\text{avg}}$ (ppm)	$\Delta\Delta\delta_{\text{avg}}$ MTS-A81C and WT (ppm)
36	0.133	0.078	-0.055
39	0.140	0.108	-0.032
40	0.040	0.019	-0.020
49	0.023	0.005	-0.018
109	0.034	0.013	-0.021
110	0.036	0.010	-0.027
119	0.015	0.036	0.021
120	0.077	0.009	-0.068
123	0.039	0.008	-0.021
176	0.017	0.078	0.061
193	0.022	0.039	0.017
195	0.011	0.050	0.039
196	0.037	0.060	0.023
197	0.135	0.006	-0.129
198	0.063	0.014	-0.049
199	0.093	0.063	-0.030
206	0.040	0.006	-0.034
207	0.032	0.012	-0.020
227	0.033	0.012	-0.021
237	0.023	0.001	-0.022
253	0.031	0.006	-0.025
257	0.047	0.028	-0.020
259	0.030	0.006	-0.025
286	0.029	0.007	-0.023
287	0.035	0.011	-0.024
291	0.060	0.026	-0.034
293	0.052	0.034	-0.018
294	0.032	0.006	-0.026

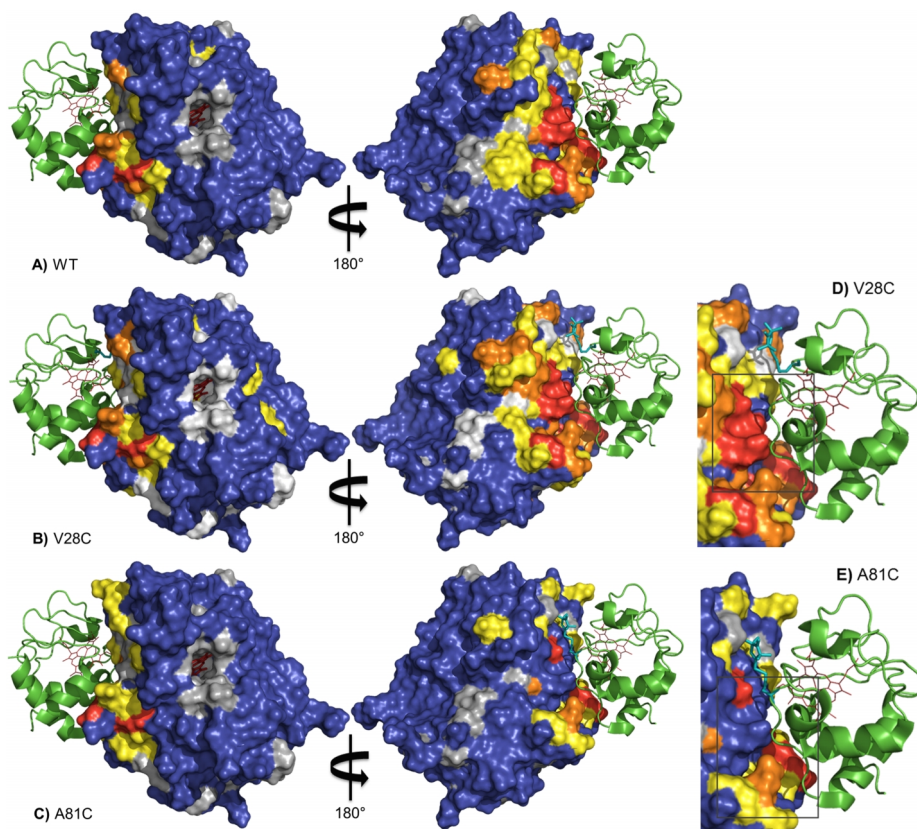


FIGURE 4.5 Chemical shift perturbation map for $^{15}\text{N}^2\text{H}$ -CcP C128A bound to WT **(A)**, MTS-V28C Cc **(B)** or MTS-V28C Cc **(C)** colour coded on a surface model of CcP (haem group in red sticks) in the stereo-specific complex (PDB-entry 2GB8).⁽¹⁾ Cc is shown in green ribbons with the haem group in red lines and MTSL rotamer A is shown in black sticks. CSP were extrapolated to 100% bound CcP. Residues with $\Delta\delta_{\text{avg}} \geq 0.06$ ppm are red, 0.04-0.06 ppm are orange, 0.02-0.04 ppm are yellow, 0-0.02 ppm are blue and with no data are grey. Residues with a large increase in $\Delta\delta_{\text{avg}}$ for MTS-Cc compared to WT are located in the black boxes for V28C **(D)** and for A81C **(E)**.

For both WT and MTS-V28C Cc, the CSP effects on CcP were localized around the binding interface where Cc is expected to be in the stereo-specific complex. The few peaks with significant differences in $\Delta\delta_{\text{avg}}$ for MTS-V28C (table 4.1) are in the centre of the binding interface (square box in D in Figure 4.5). However, overall the differences between the two CSP maps are small and the K_{B} values (Figure 4.3) are the same within error, indicating that the effects of the tag on complex formation are minimal. For A81C-V28C Cc, the CSP effects were located around the binding interface as expected but their magnitude was clearly lower than those for WT and MTS-V28C Cc. This is consistent with a higher proportion of encounter complex as suggested by the lower binding constant. It has been established that

CSP in the encounter complex are much smaller than in the specific complex.^(9, 27, 284) The peaks with a significant difference in $\Delta\delta_{\text{avg}}$ for MTS-A81C compared to WT (table 4.2) are located in the centre of the binding interface (square box in E in Figure 4.5) as is also the case for the similarly affected peaks of MTS-V28C; however, in this case the $\Delta\delta_{\text{avg}}$ is slightly decreased compared to WT.

PARAMAGNETIC RELAXATION ENHANCEMENT

While CSP analysis can be used to determine how complexes interact and even provide restraints for modelling, PRE effects are much more sensitive to weak, transient interactions and lowly populated states due to their strong distance dependence (r^{-6}). This makes them much more suitable for studying encounter complexes. A PRE map was generated on the surface of CcP of PRE caused by MTSL-V28C or MTSL-A81C Cc (Figure 4.6).

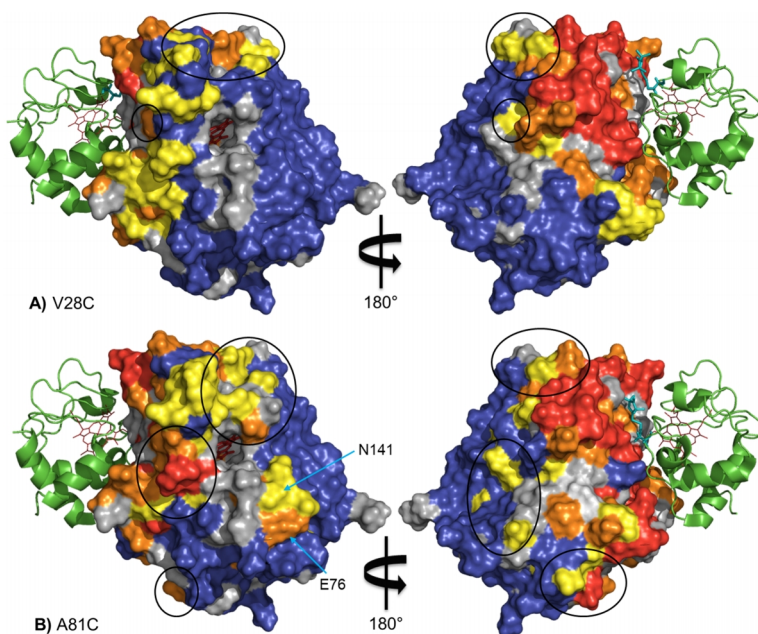


FIGURE 4.6 PRE map for $^{15}\text{N}^2\text{H}$ -CcP C128A in the presence of MTSL labelled Cc colour-coded on a surface model of CcP (haem group in red sticks) in the complex (PDB-entry 2GB8).⁽¹⁾ Cc is shown in green ribbons with the haem group in red lines and MTSL rotamer A is shown in teal sticks. **(A)** PRE map for MTSL-V28C Cc. Residues 167, 188, 213, 230-243, 245, 247, 265 and 269 are located in the black circles. **(B)** PRE map for MTSL-A81C Cc. Residues 75-79, 99-111, 141-143, 148-151, 164, 167, 182-191, 216-218, 230-243, 269 and 271 are located in the black circles. The experimental PRE was measured in a sample in which 72% (V28C) or 81% (A81C) of CcP was bound and the PRE data were then extrapolated to 100% bound CcP for this map. Residues with $\Gamma_{2,\text{para}} \geq 100 \text{ s}^{-1}$ are red, $20 \text{ s}^{-1} < \Gamma_{2,\text{para}} < 100 \text{ s}^{-1}$ are orange, $5 \text{ s}^{-1} < \Gamma_{2,\text{para}} < 20 \text{ s}^{-1}$ are yellow, $\Gamma_{2,\text{para}} \leq 5 \text{ s}^{-1}$ are blue and with no data are grey.

For both mutants, the strongest PRE effects were localized to the stereo-specific binding interface, which is consistent with the CSP map (Figure 4.5), but now the strongest effects (shown in red in both Figures) are localized slightly differently. In the CSP map, the strongest interactions occur in a large patch in the bottom half of the binding interface, while this is shifted to a smaller patch at the top corner of the binding interface in the PRE map, near the locations of the SLs at V28C or A81C. This difference is due to the different types of effects being observed; the CSP map shows the strongest interactions where the amide groups feel the strongest perturbation in their chemical environment while the strongest PRE effects occur close to the SLs. Despite this slight difference in how the effects were focused, the majority of both types of effects were localized in the same area around the binding interface. The PRE effects however formed a much larger circumference around the interface of the stereo-specific complex. This demonstrates clearly how much more sensitive PRE is for weak interactions and how they complement CSP data.

The PRE effects were converted to distances between affected residues and the paramagnetic centre. Previously, pNMR studies on the Cc-CcP complex demonstrated that 30% of the complex population was in the encounter state,^(1, 8) so the experimental data were expected to best match the predicted data for such a complex. For the PRE calculations, an estimate of 16 ns was used for the effective τ_c , the correlation time for the SL-to-nucleus vector, which incorporates contributions from at least three types of mobility. First, there is the rotational diffusion of the entire Cc-CcP complex. Second, Cc rotates within the complex relative to CcP. The τ_c for Cc movement within the encounter complex has never been determined, so it is unknown how much it contributes to the overall τ_c . Third, the rotation of the SL may contribute to the correlation time. This contribution is dependent on the distance between the SL and nucleus. The further the nucleus, the smaller the rotation angle of the SL appears to be. Sixteen ns was used because it has been demonstrated before that this value gave a good fit to the experimental, PRE derived distances, suggesting that the overall rotation of the complex dominates τ_c .⁽¹⁾ Because the orientation of the SL in the complex is unknown and may vary, several widely spaced rotamers were used (Figure 4.2) to back-calculate distances for six or twelve 30% encounter/70% stereo-specific data sets, the average of which (\pm two standard deviations) was compared to the experimental data (Figure 4.7).

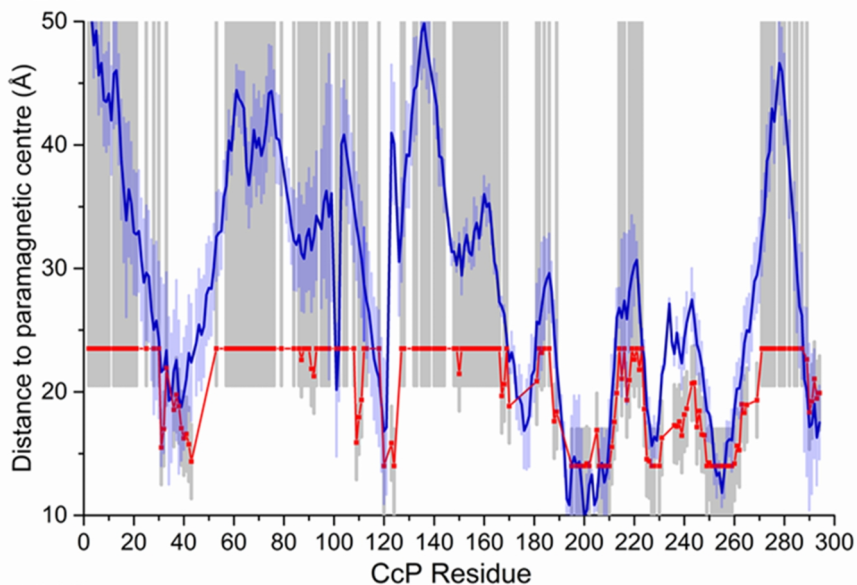
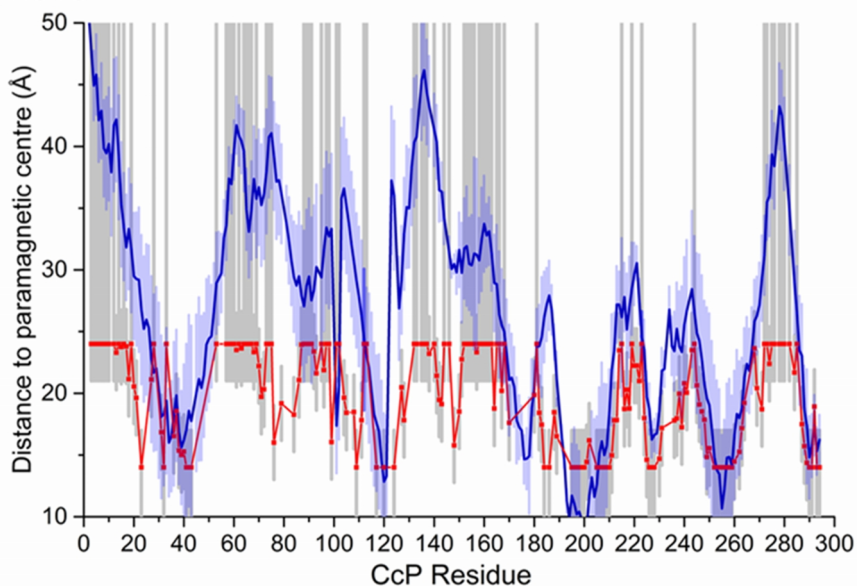
A) V28C**B) A81C**

FIGURE 4.7 Experimental and back-calculated averaged distances between CcP C128A amide protons and the paramagnetic centre in MTSL-V28C Cc (**A**) or MTSL-A81C (**B**) plotted against the CcP residue number. The red line represents the experimental distances with errors in grey bars. The average of back-calculated distances for a 30% encounter/70% stereo-specific complex for spin label rotamers A-D is shown as a blue line with a spread of two standard deviations shown in light blue bars. The experimental data were obtained with 73% CcP bound to MTSL-V28C Cc or 81% CcP bound to MTSL-A81C Cc and extrapolated to 100%.

The PRE derived distances could be determined accurately between 14-24 Å ($\pm \geq 3$ Å), and in this range there was a good global agreement between the experimental and predicted distances for both Cc mutants. However, despite the large margins for error, there were significant differences between the predicted and experimental distances for several residues. For the complex with MTSL-V28C, these were 43, 101, 102, 109, 167, 188, 205, 230-243, 245, 247, 265 and 269; for MTSL-A81C these were 23, 41-43, 75-79, 99-111, 141-143, 148-151, 164, 167, 182-191, 216-218, 230-243, 269 and 271. Furthermore, for most residues, MTSL-A81C generated stronger/additional effects compared to MTSL-V28C, which was consistent with a higher proportion of encounter complex for MTSL-A81C. This was also observed in the PRE map (Figure 4.6B). However, without further data it is not possible to determine how much the ratio of the encounter to specific state has changed.

As mentioned above, to account for the flexibility of the SL in the distance calculations, several very different rotamers were used to ensure sufficient margins of error were generated. The actual SL orientations in the complex are unknown, so it cannot be excluded that distances for residues that are just outside the error margins are attributed to other SL orientations. Residues 41-43 and 205 are directly in the binding interface of the stereo-specific complex, and therefore sensitive to the SL orientation.

The remaining residues that were more affected than predicted are in regions that border the binding interface (black circles in Figure 4.6) or slightly towards the back of CcP (blue arrows in Figure 4.6). Some of these also showed weak effects in the CSP map for WT (residues 99, 101, 102, 106, 108, 109, 110, 188, 230, 231, 236, 237), MTS-V28C Cc (residues 99, 101, 102, 107-109, 143, 164, 188, 230, 235, 236) and MTS-A81C Cc (residues 108, 167, 188, 230) (Figure 4.5). Similar effects were observed in the PRE data for 30% bound CcP for both mutants (not shown). The predicted data is based on a theoretical encounter complex simulation that was generated using an electrostatics-based MC method.⁽⁸⁾ Although it is a good representation, this model does not perfectly describe the encounter complex ensemble.⁽³⁵⁾ The observed discrepancies are relatively minor but significant, indicating that a larger data set will allow for better refinement of the model.

Interestingly, these effects were not observed by Bashir *et al.* in 2010 when they placed MTSL at ten locations on the surface of CcP and observed the PRE on Cc. In particular, MTSL was attached to L213C and S263C, which are located on either side of the region bordering the binding interface where we observe effects, but few effects were observed and none stronger than $I_{\text{para}}/I_{\text{dia}} = 0.8$. In this study MTSL was also attached to three residues close to I102 and E76 (V10, K97C and T137). MTSL at C137 showed weak effects ($I_{\text{para}}/I_{\text{dia}} =$

0.8-1.0) for most Cc residues, as did MTSL at C10 around Cc residue 20, but MTSL at C97, which is located closest to 102 and E76, did not cause any effects on Cc. ⁽⁸⁾ It could be that the presence of MTSL at these locations interfered with encounter complex interactions at that site, resulting in weak/no observed PRE. It should be noted that the same holds true for this work; even though the CSP map and the affinity were hardly affected by the MTSL at V28C, the binding was clearly affected by MTSL at A81C, which likely also increased the proportion of the population in the encounter state.

This work highlights the importance of obtaining a comprehensive data set, by using paramagnetic tags located at several sites on both sides of the complex, in order to achieve a full understanding of how the proteins interact. This is especially the case when working with ensemble data sets such as those produced by PRE. Another consideration is the flexibility of MTSL. Although MTSL tags are great tools for mapping surface interactions, their inherent flexibility limits the precision of the data, so other, more rigid tags may be more useful for refinement of the encounter complex. Expanding these studies to include data from both sides of the complex and with different types of paramagnetic tags will thus allow for a more complete characterization of the Cc-CcP complex, including refining the encounter ensemble and possibly validating the proposed low-affinity binding site.⁽¹²⁰⁾

The cytochrome c peroxidase and cytochrome c encounter complex: The other side of the story

CHAPTER 5

Protein docking using an ensemble of spin labels optimized by intra-molecular paramagnetic relaxation enhancement

Based on the research article: **Schilder J.**, Liu W.M., Kumar P., Overhand M., Huber M. and Ubbink M. (2015). *Protein docking using an ensemble of spin labels optimized by intra-molecular paramagnetic relaxation enhancement*. Phys Chem Chem Phys., DOI: 10.1039/c5cp03781f.

ABSTRACT

Paramagnetic NMR is a useful technique to study proteins and protein complexes and the use of PRE for this purpose has become wide-spread. PRE are commonly generated using paramagnetic SLs that contain an unpaired electron in the form of a nitroxide radical, with MTSL being the most popular tag. The inherent flexibility of the SL causes sampling of several conformations in solution, which can be problematic as over- or underestimation of the spatial distribution of the unpaired electron in structural calculations will lead to errors in the distance restraints. We investigated the effect of this mobility on the accuracy of protein-protein docking calculations using intermolecular PRE data by comparing MTSL and the less mobile pyMTSL on the dynamic Cc-CcP complex. No significant differences were found between the two SLs. Docking was performed using either single or multiple conformers and either fixed or flexible SLs. It was found that mobility of the SLs is the limiting factor for obtaining accurate solutions. Optimization of SL conformer orientations using intra-molecular PRE improves the accuracy of docking.

INTRODUCTION

Paramagnetic NMR is a convenient approach for determining the binding site and orientation of proteins within low-affinity complexes that undergo minimal structural changes upon binding. Various types of paramagnetic restraints can be used, such as PCS, RDC induced by the alignment caused by the paramagnetic centre, or PRE.^(58, 59) PRE is the most popular choice due to the simplicity of introducing relaxation centres to proteins, mainly in the form of site-specific tags, of which SLs are the most common. SLs are small organic compounds that contain an unpaired electron in the form of a nitroxide radical and are generally quite stable under non-reducing conditions.⁽²⁸⁵⁾ The electron spin can be observed directly using electron paramagnetic resonance (EPR) or indirectly by measuring the PRE effects on nearby nuclei via NMR spectroscopy. The observed nuclear relaxation rates can either be used directly or after conversion into distances for structural modelling.^(63, 77)

To convert the observed relaxation rates into distances, the correlation time (τ_c) of the vector \mathbf{r} that connects the paramagnetic centre and the nucleus is required. τ_c depends of the rotational correlation time of this vector (τ_r) as well as the longitudinal electronic relaxation time (τ_s), according to $\tau_c^{-1} = \tau_r^{-1} + \tau_s^{-1}$.^(63, 77) For SLs, the contribution of τ_s is small and τ_c is dominated by τ_r . Any motions that change the vector will have an effect on τ_c , such as protein tumbling, SL mobility, local protein dynamics, and, for intermolecular PRE measured in a complex, the motions of one protein relative to the other.⁽²⁸⁶⁾ Thus, it is not straight forward to determine τ_c but, fortunately, due to the sixth power dependence of the PRE on the distance r between SL and the nucleus, errors in τ_c result in only small errors in the distances.^(89, 287)

The differences between the free energies of SL conformations are often smaller than the thermal energy in the sample, meaning that the position of the unpaired electron is spatially distributed over an area determined by the occupied SL conformer orientations.⁽²⁸⁶⁾ Over- or underestimation of the spatial distribution of the unpaired electron will lead to errors in the apparent mean distance, [$\langle r^{-6} \rangle^{-1/6}$].⁽⁸⁹⁾ To solve this problem, the SL can be treated as an ensemble of non-self-interacting conformers during simulated annealing calculations.⁽⁸⁹⁾ In this case, the number of conformers required to make up the ensemble depends on the actual spatial distribution of the SL in solution as well as the precision required to match the experimental data for nearby nuclei, which will experience the strongest PRE. However, using an ensemble could potentially generate worse results if non-realistic conformer orientations are used, leading to an inaccurate structure of the protein complex.⁽²⁸⁶⁾ This can be overcome by first determining the most favourable conformer orientations experimentally using intra-

molecular PRE data, followed by fixing the SL conformers in those positions during docking calculations based on inter-molecular PRE data. This has yielded good results using as few as one SL conformer for multi-domain proteins^(287, 288) and has also been used to study DNA-protein interactions.^(28, 289, 290)

Due to its wide commercial availability, as well as the availability of a suitable diamagnetic control, the most commonly used SL is MTSL. However, MTSL has a rather long linker consisting of five single bonds from the peptide C α atom to the 3-pyrroline ring (Figure 5.1B), resulting in substantial flexibility and dynamics.

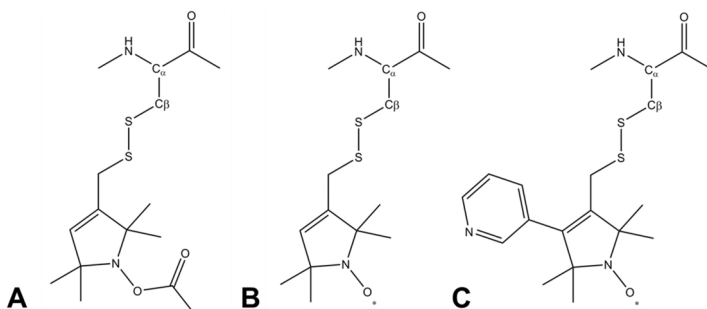


FIGURE 5.1 Molecular structures of three methanethiosulfonate based tags attached to a cysteine side chain. **(A)** MTS (diamagnetic control), **(B)** MTSL and **(C)** pyMTSL.⁽²⁹¹⁾

In order to limit tag dynamics, the pyrroline ring of MTSL can be modified to increase the tag rigidity.⁽²⁸⁶⁾ This was successfully done via the addition of a pyridyl group to position 4 of the pyrroline ring resulting in pyMTSL (also called HO-3606) (Figure 5.1C). It was reported that this addition restricts the SL movement, such that the PRE data could be fit using just one conformer.⁽²⁹¹⁾

The Cc-CcP complex is a well-studied, highly dynamic ET complex. CcP catalyzes the reduction of hydrogen peroxide to water using two electrons it receives from Cc.⁽¹³⁵⁾ The crystal structure was determined in 1992, showing how Cc is docked on CcP.⁽¹⁶⁹⁾ The orientation of Cc relative to CcP in the complex in solution was determined by Volkov *et al.* in 2006 by attaching MTSL at five positions on CcP. After confirming that the presence of the SLs did not interfere with complex formation, the PRE effects were measured for nuclei in Cc. The orientation of Cc was determined using rigid body docking with an ensemble of four MTSL conformers per position in orientations selected to represent the width of the ensemble of sterically allowed conformers, yielding a position of Cc that was close to the one observed in the crystal structure (RMSD=2.2 Å for the C α atoms of Cc).⁽¹⁾ The data also provided evidence

for the presence of an encounter complex, in which Cc assumed other orientations relative to CcP. In 2010 Bashir *et al.*, described similar rigid body docking calculations in which the MTSL conformers at each position were treated as an ensemble of non-self-interacting conformers that could move freely during the docking, as could the nearby amino acid side chains. The result was a less precise protein complex ensemble, in which extreme positions for Cc were found within the ensemble when all the MTSL conformers were simultaneously orientated to one side or another; this effect appeared to have been averaged out in the previous study when the conformers were fixed in four dispersed orientations.⁽¹³²⁾

In this work, we revisit the Cc-CcP complex using MTSL and pyMTSL tags at three positions close to the binding interface on the surface of CcP. The aim of this work is to establish whether PRE data alone are sufficient for accurate rigid body docking of two proteins that form a dynamic complex. The Cc-CcP complex spends approximately 30% of the time in an encounter state;^(1, 8, 35, 140, 204) therefore, the PRE represent not only the well-defined stereospecific state but also the encounter state. We investigate the role of SL mobility by comparing MTSL to pyMTSL, as well as by comparing the use of a single SL conformer to the use of an ensemble of conformers when the orientations are either fixed or mobile during the calculations. The position of Cc in the crystal structure has been used as the benchmark, under the assumption that the well-defined state of the complex in solution is very similar to the one observed in the crystalline state. It is concluded that the results of PRE based docking are highly dependent on the choice of SL conformers. Thus, the SL flexibility appears to be the limiting factor for obtaining accurate results. SL ensembles of conformers optimized by intramolecular PRE yield the best results.

MATERIALS & METHODS

PROTEIN PURIFICATION

The genes for the yeast CcP C128A containing mutations NC38, NC200 and TC288⁽¹⁾ were sub-cloned in the pET28aCcP plasmid and were expressed to produce CcP, which was purified as described in previously (Chapter 3).⁽¹⁴⁰⁾ The yields were 40 mg/L for NC38 [²H¹⁵N], 20 mg/L for NC200 [¹⁵N] and 120 mg/L for TC288 [²H¹⁵N] in minimal media. A pUC19 based plasmid containing the *S. cerevisiae* iso-1-cytochrome *c* gene was used to produce Cc, which was purified according to published procedures.^(124, 125) The yield was 20 mg/L [¹⁵N] in minimal media. The concentrations of CcP and Cc were determined using UV-Vis spectroscopy with $\epsilon_{408\text{nm}} = 98 \text{ mM}^{-1}\text{cm}^{-1}$ and $\epsilon_{410\text{nm}} = 106.1 \text{ mM}^{-1}\text{cm}^{-1}$, respectively.^(125, 220)

SPIN LABEL PREPARATION

MTS and MTSL tags were obtained from Toronto Research Chemicals (North York, ON, Canada) and pyMTSL was synthesized according to the published protocol by Dr. Weimin Liu.⁽²⁹¹⁾ The SLs were stored as 100 mM stocks dissolved in DMSO at 4 °C prior to use. The CcP mutants were tagged with either MTS, MTSL or pyMTSL, as described previously (Chapter 4).^(1, 140) The tagging efficiency was determined by mass spectroscopy to be essentially 100% and SLs at these positions have previously been shown not to interfere with Cc-CcP complex formation.⁽¹⁾

CONTINUOUS WAVE-EPR EXPERIMENTS

The SL mobility was determined using X-band continuous wave (cw) EPR measurements. These were performed using an ELEXSYS E680 spectrometer (Bruker, Rheinstetten, Germany) with a rectangular cavity. All the measurements were done at room temperature (20 °C), using 0.6346 mW (0.5 mW for N38C) microwave power, 100 kHz modulation frequency and 0.08 mT (0.05 mT for N38C) modulation amplitude. The total measurement time was between 1 and 2 hrs per spectrum.

The cw-EPR spectra were simulated using Matlab version 7.14.0.739 (Natick, Massachusetts, U.S.A) and the EasySpin package version 4.5.5.⁽²⁹²⁾ For all simulations, the following spectral parameters were used: $g = [2.00906, 2.00687, 2.00300]$ ⁽²⁹³⁾ and the hyperfine tensor parameters $A_{XX} = A_{YY} = 13$ MHz. Isotropic rotation was assumed in all cases, to reduce the number of simulation parameters and get a better overall picture of the relative changes in rotational properties of the SLs. Usually a superposition of more than one component was required to simulate the spectra. One of these components has a rotation correlation time of several tens of ps and a contribution of less than 0.5 % to the total simulation and is therefore assigned to a small fraction of residual free SL in the sample.

PARAMAGNETIC NMR SPECTROSCOPY

For the intra-molecular PRE measurements, NMR samples contained 300 μ M unlabelled Cc WT and 300 μ M double labelled [¹⁵N, ²H] NC38 or TC288 or ¹⁵N-labelled NC200 CcP with either MTS, MTSL or pyMTSL tags attached in 20 mM NaPi, 100 mM NaCl, 6% D₂O, pH 6.0. 2D BEST-TROSY-HSQC experiments⁽²⁷⁹⁾ were recorded on a Bruker AVIII HD spectrometer equipped with a ¹H[¹³C/¹⁵N] TCI-cryoprobe operating at a Larmor frequency of 850 MHz at 293 K with 1024 and 100 complex points in the ¹H and ¹⁵N

dimensions, respectively. For the inter-molecular PRE measurements, NMR samples contained 300 μM ^{15}N -labelled Cc WT and 300 μM unlabelled NC38, NC200 or TC288 CcP with either MTS, MTSL or pyMTSL tags attached in the same buffer solution. 2D HSQC experiments were recorded on the same spectrometer and at the same temperature with 512 and 64 complex points in the ^1H and ^{15}N dimensions respectively. All data were processed using Topspin 3.2 (Bruker, Karlsruhe, Germany) and analyzed using CCPN Analysis 2.1.5.⁽²⁴²⁾

The ^1H and ^{15}N resonance assignments were obtained from previous studies for Cc^(183, 294, 295) and from our previous work for CcP (Chapter 3).⁽¹⁴⁰⁾ In the early stages of this work for the intra-molecular PRE of TC288-(py)MTSL, we noticed that residue 120 experienced a very strong PRE (the paramagnetic peak disappeared from the spectra) while residues 123 and 124 did not experience any PRE. It was concluded that residues 123 and 124 had been swapped with those of residues 101 and 102. BMRB entry 19884 containing these backbone resonance assignments has been updated.⁽²⁵⁵⁾

DATA ANALYSIS

The intensity ratio of the amide resonances in the spectra of the paramagnetic (MTSL- or pyMTSL-tagged) and diamagnetic (MTS-tagged) samples ($I_{\text{para}}/I_{\text{dia}}$) was measured and normalized as described previously⁽⁸⁾ The paramagnetic contribution to the transverse relaxation rate, $R_{2,\text{para}}$, was calculated as described previously (Chapter 4).^(8, 77, 140) The average $R_{2,\text{dia}}$ value was used with a large error margin for those amides for which an $I_{\text{para}}/I_{\text{dia}}$ could be measured but for which the line width of the diamagnetic peak could be not obtained. For the amide peaks that disappeared in the paramagnetic spectrum, an upper limit for I_{para} was set to two standard deviations of the noise level of the spectrum (Chapter 4).⁽¹⁴⁰⁾

The calculated $R_{2,\text{para}}$ values were then converted into distances as described previously (Chapter 4),⁽⁸⁾ see equation 4.1, with a fraction bound of 0.88 for Cc bound to CcP; 1.0 for intramolecular PRE in CcP. The value of τ_c for the complex was previously estimated to be 16 ns^(1, 8) and tests done using larger values (20 ns or 29 ns) did not improve the results.

OPTIMIZATION OF THE SL ORIENTATIONS

The protein coordinates for CcP were taken from 1ZBY for CcP.⁽¹⁴⁸⁾ The addition of surface cysteine mutations and the introduction of the SLs to the CcP structure were done *in silico* as described previously.⁽¹⁾ The initial positions for the four conformers of MTSL were generated by systematic rotation of the SL around the five single bonds that join the pyrroline

ring to the C α atom of the cysteine residue and choosing four of the sterically allowed orientations for each mutant that represented the ensemble well, as described previously.⁽¹⁾ The same orientations as reported in that article were used here. For pyMTSL, the four conformers were snapshots of short molecular dynamics runs of the SL *in vacuo*. Using the intra-molecular PRE data, a set of distance restraints was calculated and used to determine the most favourable SL conformer orientations for a single conformer or an ensemble of four conformers at each mutant position using Xplor-NIH version 2.34.^(282, 283) For localizing the SL orientations on CcP, only PRE from amides within 23 Å of the SL oxygen atom were used, thus eliminating the data that do not yield relevant restraints (class III PRE). When an ensemble of four SL conformers was used, r^{-6} ensemble averaging was done for the distances r between the four SL positions and a nucleus. The calculations were performed in two steps using side-chain dynamics, in which all atoms were fixed apart from those of the SL and the amino side chains within 10 Å of the SL. During the first step, only vdW forces were considered, followed by a second step in which both the vdW forces and the distance restraints were used. The vdW forces were defined as repel forces between the SL and the protein atoms but were set to zero between multiple SLs in the ensemble. This was repeated 1,000 times and the orientation of the lowest energy was used as the experimentally determined orientation in subsequent docking calculations of Cc to CcP.

PROTEIN DOCKING

The protein coordinates for the individual proteins were obtained from PDB 2YCC for Cc⁽¹²⁷⁾ and 1ZBY for CcP⁽¹⁴⁸⁾ and for the complex from PDB 2PCC.⁽¹⁶⁹⁾ The docking of Cc to CcP was driven by a set of distance restraints derived from inter-molecular PRE data using Xplor-NIH version 2.34.^(282, 283) This was done using either a single SL conformer or an ensemble of four SL conformers, the orientations of which were either fixed in selected positions, fixed in experimentally determined positions as described above, or were free to move during docking. Cc was docked to CcP using rigid body dynamics with vdW repel forces and the distance restraints contributing to the total energy. The vdW forces were set to zero for interactions between atoms of multiple SLs within the ensemble. Docking was repeated 10,000 times and the twenty lowest energy structures were used to generate the protein complex ensemble.

The precision of the ensemble was quantified by the average of the pairwise RMSD values of members of the ensemble and its mean structure. The accuracy was assessed by comparing the mean structure of the calculated structure ensemble to the crystal structure of

the complex, using both the root mean squared deviation (RMSD) and distance based root mean squared deviation (DRMS) for these two structures, which is defined as (see list of symbols at beginning of thesis for all symbol descriptions):^(3, 8)

$$DRMS = \frac{1}{N} \sum_{i,j} |d_{ij}^{ens} - d_{ij}^{xray}| \quad 5.1$$

The fit between the observed (dis^{obs}) and back-calculated (dis^{calc}) distances for the class II restraints was evaluated using Q-factors according to:⁽⁸⁾

$$Q = \sqrt{\frac{\sum_i (dis_i^{obs} - dis_i^{calc})^2}{\sum_i (dis_i^{obs} + dis_i^{calc})^2}} \quad 5.2$$

Note that in this definition, the denominator is the sum of the observed and calculated distances. The average violations (AV) were determined by averaging the difference between the experimental and back-calculated distances; for distances with only an upper (class I) or lower boundary (class III), back-calculated distances that fell inside of those boundaries were not considered violations.

RESULTS & DISCUSSION

SL MOBILITY STUDIED BY EPR

The inherent flexibility of SL linkers allows a SL such as MTSL to occupy several conformer orientations over time,⁽²⁸⁶⁾ which can be problematic when trying to accurately determine $R_{2,para}$ values.⁽⁸⁹⁾ Substitution of a bulky side group on the pyrroline ring was shown to restrict the movement of pyMTSL and reduce the number of allowed orientations.⁽²⁹¹⁾ In order to compare the mobility of MTSL and pyMTSL, these SLs were attached to three positions on the surface of CcP, positions 38, 200 and 288 (Figure 5.2).

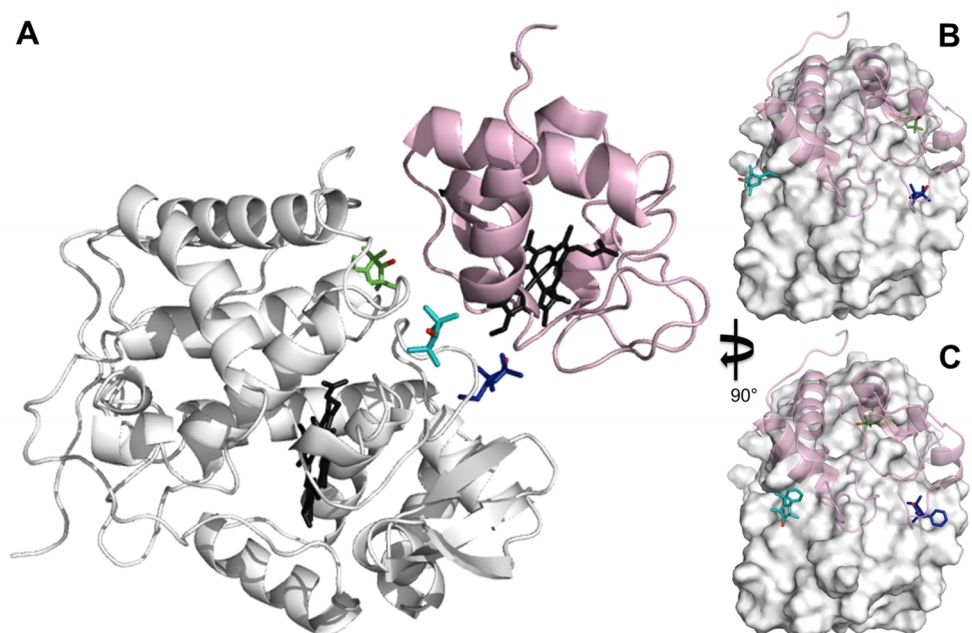


FIGURE 5.2 Locations of SLs attached on the surface of CcP at positions 38 (teal), 200 (blue) and 288 (green) (nitroxide oxygen atom in red). **A)** The locations of three MTSL tags shown on CcP (white) with respect to the binding site of Cc. The relative orientations of MTSL (**B**) or pyMTSL (**C**) around the binding site on the surface of CcP are also shown; Cc is shown as a pink cartoon shadow. The haem groups are shown in black sticks (PDB-entry 2PCC).⁽¹⁶⁹⁾

These three attachment positions ring the stereo-specific binding interface with Cc and have also previously been used to produce significant inter-molecular PRE effects on Cc.^(1, 8) The spin-label mobility at these positions for MTSL and pyMTSL was compared using EPR measured in solution at room temperature (Figure 5.3). The EPR spectra of the SLs at the three positions investigated show that the line shapes of the nitroxides are under conditions of not fully averaged anisotropy, typical for spin-labels attached to proteins. In all cases, the spectrum of the protein labelled with MTSL at a given position has narrower lines than that of the protein labelled with pyMTSL, showing that the rotation of the nitroxide group in pyMTSL is slower than that of MTSL. More details can be seen from the parameters of the spectral simulation (Figure 5.4; Table 5.1).

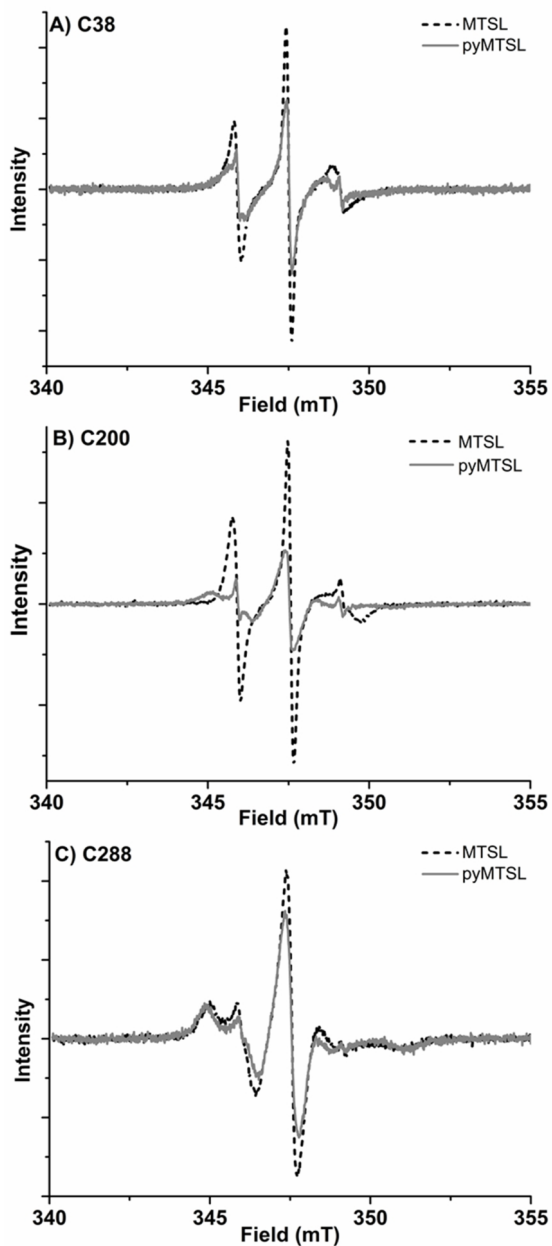


FIGURE 5.3 EPR spectra for MTSL (dashed black) and pyMTSL (solid grey) at positions 38 **(A)**, 200 **(B)** and 288 **(C)** on CcP. Spectra are normalized to the number of spins, which emphasises differences in line width; narrow lines have larger amplitudes.

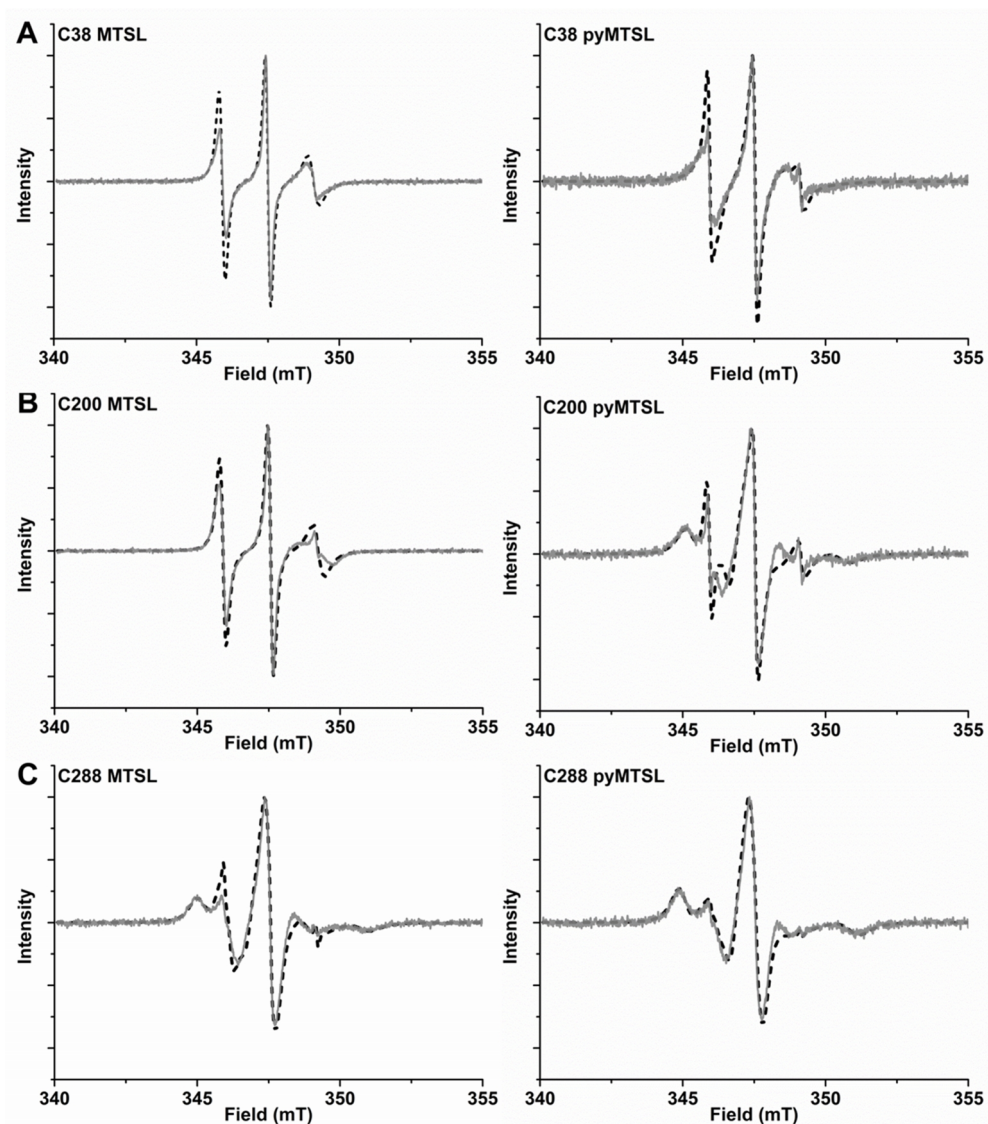


FIGURE 5.4 Experimental spectra (solid grey) and corresponding spectral simulations (dashed black) of MTSL and pyMTSL at positions 38 (A), 200 (B) and 288 (C) on CcP. Simulation parameters, see Table 5.1.

TABLE 5.1 Parameters of the simulations of the EPR spectra of spin labelled CcP mutants.

	Component 1			Component 2			Free MTSL*		
	τ_r^{radical}	Fraction	A _{zz}	τ_r^{radical}	Fraction	A _{zz}	τ_r^{radical}	Fraction	A _{zz}
	(ns)	(%)	(MHz)	(ns)	(%)	(MHz)	(ns)	(%)	(MHz)
C38 MTSL	0.89	100.0	110.0	n/a	n/a	n/a	n/a	n/a	n/a
C38 pyMTSL	0.63	20.0	108.5	2.20	79.5	102	0.05	0.5	108.0
C200 MTSL	0.89	99.7	115.0	n/a	n/a	n/a	0.02	0.3	109.8
C200 pyMTSL	0.81	17.0	109.8	8.70	82.7	94	0.05	0.3	108.0
C288 MTSL	2.23	42.0	100.0	10.00	57.7	97	0.02	0.3	109.8
C288 pyMTSL	3.20	25.0	101.0	10.00	74.9	101	0.05	0.1	108.0

* residual free spin-label fraction, contribution shows almost complete removal of free spin label from all samples

For CcP positions 38 and 200, MTSL has a rotation correlation time (τ_r^{radical}) of 0.9 ns, which is the single component contributing to these spectra. When using the pyMTSL tag, the radical has a majority contribution with a 2.5 fold (position 38C) and a more than ten-fold increased rotation correlation time (position 200C) (see Table 5.1). For the 288 position the MTSL label is rotating more slowly than at the other positions: two contributions are observed with almost equal weight and correlation times of 2 ns respectively 10 ns, which is significantly more immobilized than the MTSL at the other positions. Replacing MTSL by pyMTSL increases the rotation correlation time of the faster fraction to 3 ns and shifts the population to 75 % of slower fraction, showing that also at this position, the pyridine substituent leads to a slowing down of the rotation.

The rotation correlation times of pyMTSL vary per position, showing that local interactions contribute to the mobility of pyMTSL and that the mobility of pyMTSL is not exclusively determined by its side-chain structure. Furthermore, none of the rotation correlation times reach the value of the protein, showing that the SL is not completely anchored to the protein, and demonstrating that the linker connecting the nitroxide-ring to the protein backbone possesses local degrees of freedom. Note that EPR reports on the correlation time of the radical, whereas PRE reflect the rotational correlation time of the radical-nuclear vector (τ_r).

EXPERIMENTAL OPTIMIZATION OF SL CONFORMERS

We determined the conformer orientations for both MTSL and pyMTSL using intramolecular PRE to see if a difference could be observed between these tags.⁽²⁹¹⁾ In order to account for potential differences in the most favourable SL orientations on free CcP compared to CcP in complex with Cc, all intramolecular PRE measurements were done in the presence of 300 μM non-isotopically labelled Cc, which was the same concentration as used for the

intermolecular PRE measurements. The PRE were measured for the SLs at three positions of CcP and were then converted into distances between the affected nuclei and the paramagnetic centre. These distance restraints were used to determine the favourable orientations of a single conformer or an ensemble of four conformers at each position for both MTSL and pyMTSL (Figure 5.5). When the SL orientations were fit to the data using only a single conformer, the position of the nitroxide radical for the twenty lowest energy structures converged to a fairly precise position at all three attachment sites (Figure 5.5-i). Also, little difference was seen between MTSL and pyMTSL. When using an ensemble of four conformers, however, the most favourable orientations were more dispersed and two or three distinct populations appeared (Figure 5.5-ii). Table 5.2 reports the Q-values and average violations (see Materials and Methods for the definition of these measures) of the various fits. Interestingly, the ensemble fit yields results that are only marginally better than those obtained with the single conformers. Thus, on the basis of these data it cannot be decided which is the better description of SL conformations.

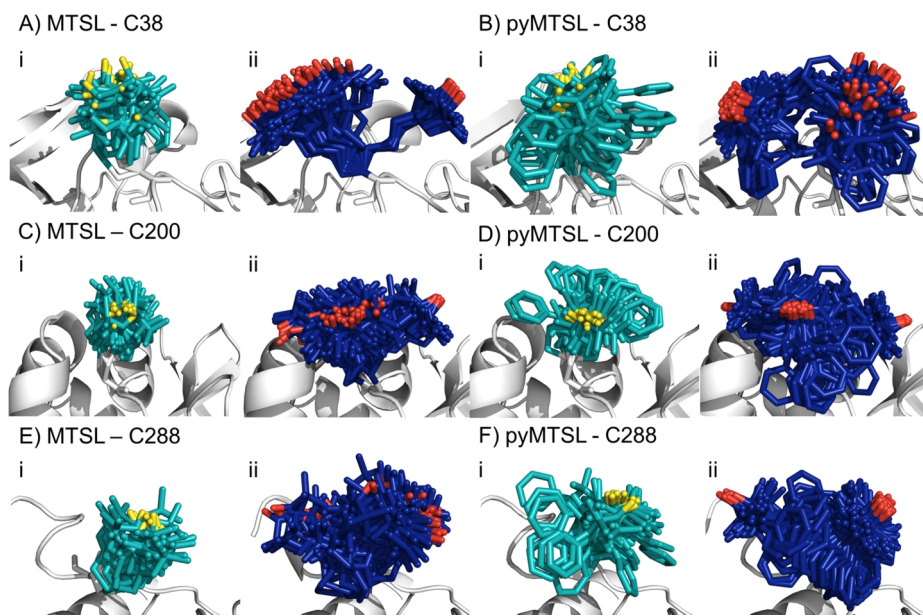


FIGURE 5.5 The twenty lowest energy conformer orientations found using intra-molecular PRE based distance restraints for MTSL attached at C38 (**A**), C200 (**C**) or C288 (**E**) or pyMTSL attached at C38 (**B**), C200 (**D**) or C288 (**F**) on the surface of CcP (grey ribbon). The PRE data were fit using a single conformer (i) or an ensemble of four conformers (ii). For a single conformer, the SLs are shown in cyan sticks and the nitroxide oxygen atom is shown in yellow. For the conformer ensemble, the SLs are shown in blue sticks and the nitroxide oxygen atom is shown in red.

Essentially the same position was found for the nitroxide radical when using either MTSL or pyMTSL. For C38-MTSL (Figure 5.5 A,B-ii), the ensemble showed two distinct populations, one with a very well-defined location of the nitroxide radical and one more dispersed, with ratio of approximately 25%:75%. For pyMTSL, the positions are similar but the populations have a ratio closer to 50%:50%. Furthermore, the well-defined and disperse populations are swapped when compared to those for MTSL. The EPR data (Figure 5.3A) showed a marked decrease in mobility for pyMTSL compared to MTSL at position C38 but this difference was not reflected in the number of calculated SL orientations. Furthermore, the populations found by NMR were not reflected in the EPR simulations (Table 5.1). For example, for C38 MTSL the EPR simulations found only one highly mobile population while for C38 pyMTSL the EPR simulations found two populations but in a ratio of 20.0%:79.5% as compared to roughly 50%:50% found by NMR. Differences between the ratios of populations found by EPR and NMR were also seen for both SLs at C200 and for MTSL at C288 (see discussion below).

For position C200, the EPR data showed a similar decrease in mobility for pyMTSL compared to MTSL, which is reflected in a much more defined positioning of the nitroxide radical for pyMTSL compared to MTSL in the calculated SL ensemble orientations, although the general location of the radicals was the same in both cases (Figure 5.5 C,D). When using an ensemble of four conformers at this position, three distinct populations were found with a ratio of about 25%:50%:25% for both SLs. However, the EPR simulations found only one population for C200 MTSL and only two populations for C200 pyMTSL with a ratio of 17.0%:82.7% (Table 5.1).

The NMR spectra for C288-(py)MTSL showed CSP for peaks of the residues near the SL attachment site (small shifts for residues 284 and 285; larger shifts for residues 286-294) indicating that the presence of the SL affects the local backbone structure of the C-terminal loop. Therefore, the position of the nitroxide radical was also determined by docking pseudoatom(s) to CcP using the intra-PRE data, representing the paramagnetic centre(s) unrestrained by the covalent linkage to the CcP backbone. The resulting positions and fit to the data were very similar to those found when the SLs were attached to CcP, indicating that linking the SL to the backbone of CcP seen in the crystal structure did not interfere with determining the experimentally most favourable orientations. As observed for the other SL positions, the ensemble of conformers at C288 showed more than one population for both SLs, with a ratio of about 25%:75%. Interestingly, the EPR simulations also showed two

populations for both SLs at C288 with a ratio of 42.0%:57.7% for MTSL and a ratio of 25.0%:74.9% of pyMTSL.

Overall, there seems to be little correlation between the populations of SL conformers found using NMR and those found during the EPR simulations; C288 pyMTSL is the only position for which ratio between the populations found with NMR matched those found with EPR. Furthermore, the τ_r^{radical} values found during the EPR simulations (0.63-3.20 ns for the fast components; 2.20-10.00 ns for the slow components) were much lower than the estimated τ_c of the protein (16 ns).^(1, 8) For SLs, the value of τ_s is large so τ_c is dominated by τ_r in the formula $\tau_c^{-1} = \tau_r^{-1} + \tau_s^{-1}$.^(63, 77, 286) The value of τ_r is affected by any motions that change the vector between the paramagnetic centre and the nucleus such as protein tumbling, SL mobility, and local protein dynamics⁽²⁸⁶⁾ but of these usually only protein tumbling is taken into account⁽²⁹⁶⁾ as it is generally assumed to dominate the relaxation measurements.⁽²⁹⁷⁾ Despite shorter τ_r^{radical} values for the SLs, the assumption that protein tumbling dominates τ_c appears to hold true here as previous tests on the same system using lower τ_c values of 4 or 12 ns did not improve the docking results.⁽¹⁾ This discrepancy between τ_r and τ_r^{radical} can be understood if it is assumed that the short correlation of the radical is caused mainly by small movements of the SL but have little influence on the length and orientation of the radical-nuclear vector in the external field. Rearrangement of this vector is dominated by rotation of the protein.

The EPR spectra for position C288, showed that both MTSL and pyMTSL were highly immobilized (Figure 5.3C), so highly defined locations for the most favourable conformer orientations may be expected. When using a single conformer, the nitroxide radical positions were indeed precisely defined (Figure 5.5E,F-i) and also the radical positions for the pyMTSL ensemble are well defined (Figure 5.5E-ii). However, it should be noted that the calculations give no evidence for stronger steric restrictions. The restraints used in the calculations are to the oxygen of the SL. Other atoms can sample the conformational space as far as allowed by the restrained position of the oxygen atom. This is illustrated by the pyridyl ring that occupies a wide range of orientations for the pyMTSL in each of three positions. It is not obvious that this range is more limited for C288.

To see how well the calculated structures of the most favourable orientations fit the PRE data, the distances from the paramagnetic centre to the amide protons were back-predicted using r^{-6} averaging over all conformer orientations present in the best twenty

solutions (Figure 5.5) and then compared to the experimentally observed distances (Figure 5.6) and the quality of fit parameters for all data sets are given in Table 5.2.

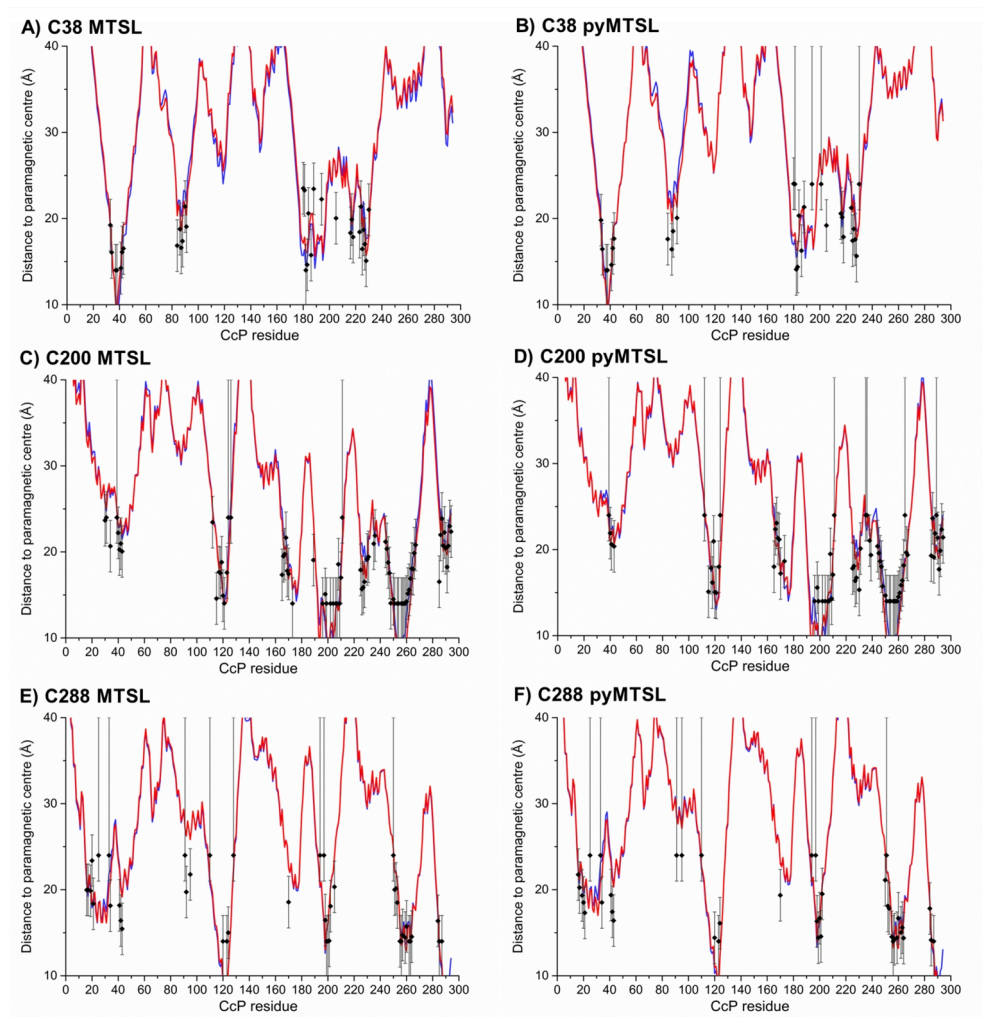


FIGURE 5.6 Experimental and back-calculated distances between CcP C128A amide protons and the paramagnetic centre in MTSL or pyMTSL at positions 38 (**A,B**), 200 (**C,D**) or 288 (**E,F**) plotted against the CcP residue number. The black diamonds represent the experimental distances with error bars. Only PRE of residues within 23 Å of the paramagnetic centre were used in the calculations. The average back-calculated distances are also shown as lines for a single SL conformer (red) or an ensemble of four conformers (blue). The PRE were measured while CcP was in a 1:1 complex with Cc.

TABLE 5.2 Q-factors and average violations (AV) for the fit of the back-calculated to the experimental distances derived from intra-molecular PRE between CcP amide protons and the paramagnetic centre in MTSL or pyMTSL at positions 38, 200 or 288 of CcP.

Position		MTSL		pyMTSL	
		1 conformer	4 conformers	1 conformer	4 conformers
C38	AV (Å)	2.94	2.44	2.83	2.57
	Q-factor	0.20	0.17	0.18	0.17
C200	AV (Å)	1.65	1.61	1.82	1.80
	Q-factor	0.15	0.14	0.15	0.14
C288	AV (Å)	2.35	2.28	2.16	2.13
	Q-factor	0.18	0.17	0.16	0.15

Very little difference is seen between the calculated distances, and thus the quality parameters, for the single conformer and ensemble solutions. Therefore, it cannot be established whether the single conformer or the ensemble is the better description for the SL, so both have been used in the docking calculations to allow for a comparison. Also little difference is present between solutions for MTSL and pyMTSL, in accord with similarity between the input PRE data sets for both SL types. There are, however, significant differences in the quality of fit between the SL positions. The best fits are observed for C200. The reason is not evident.

DOCKING OF CC TO CCP SINGLE SL CONFORMER

For the docking of Cc to CcP, inter-molecular PRE data were obtained for Cc in complex with CcP that had been spin-labelled with either MTSL or pyMTSL at positions C38, C200, or C288. The distances between Cc amide protons and SL oxygen atoms derived from the PRE were used in restrained rigid-body docking of Cc to CcP. It should be noted that the complex comprises a significant fraction of encounter complex, in which Cc is in an orientation close to but different from the stereo-specific complex, as has been demonstrated before.(14, 15, 19) The free proteins, encounter state and stereo-specific complex are all in fast exchange on the NMR timescale.(14) Thus, docking of a single Cc molecule solely based on PRE derived distance restraints is not expected to give a perfect fit to the data, because the contribution of the encounter complex to the PRE is ignored. Nevertheless, it was shown that such docking can yield a structure that is close to the crystal structure of the stereo-specific complex,(14) and, furthermore, this issue will be of relevance for many weak and transient complexes.

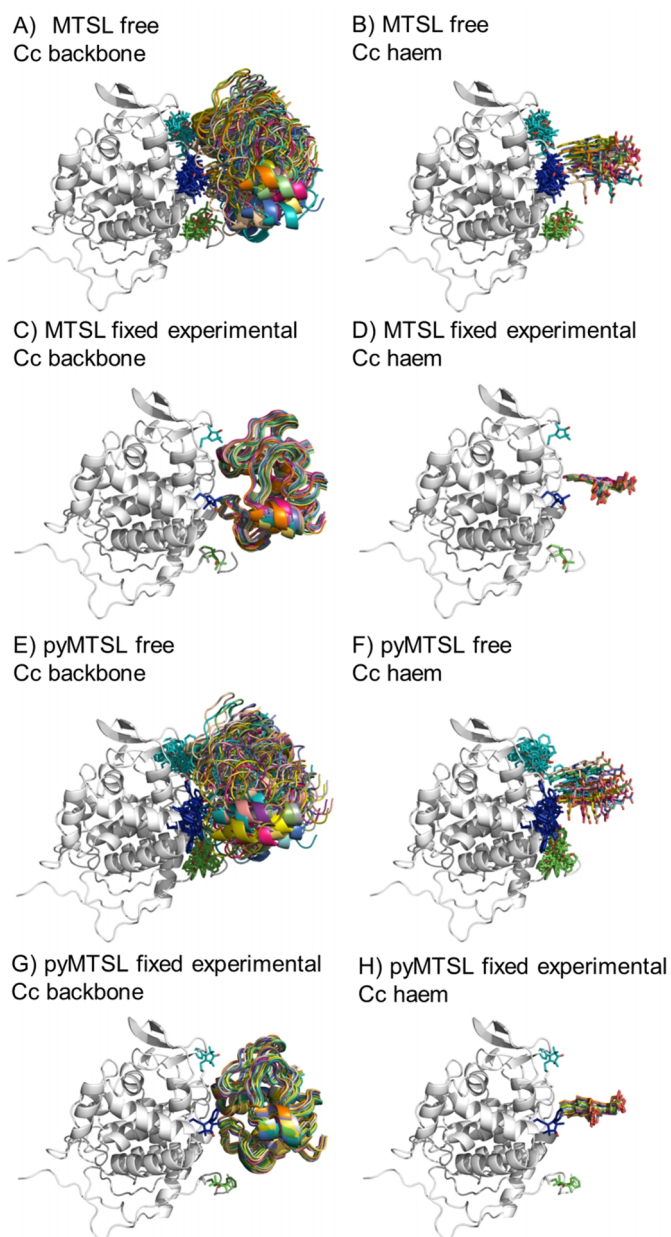


FIGURE 5.7 Twenty lowest energy solutions for docking Cc to CcP driven by intermolecular PRE data using a single SL conformer that was free to move during DD (MTSL **A, B**; pyMTSL **E, F**) or that was fixed in the experimentally determined most favourable orientation (MTSL **C, D**; pyMTSL **G, H**). CcP is shown in grey ribbons and Cc is shown in multi-coloured ribbons. The SLs are shown in sticks at positions 38 (teal), 200 (blue) and 288 (green), with the nitroxide oxygen atom in red and the Cc haem group is shown in multi-coloured sticks. The docking was done using the Cc and CcP structures taken from PDB entry 2PCC.⁽¹⁶⁹⁾

During the docking, first a single SL conformer was used that was either free to rotate or fixed in the experimentally determined most favourable orientation. Docking was repeated 10,000 times from random starting positions of Cc and an ensemble was generated from the twenty lowest energy solutions for MTSL (Figure 5.7 A-D) and pyMTSL (Figure 5.7 E-H). When the SL is free to move during docking, the position of Cc as well as the SL orientations in the resulting ensemble of the twenty lowest energy solutions were much more dispersed than when the SL orientations were fixed; this can be seen most clearly when the positions of the Cc haem groups are compared. Similar results were observed in a previous study.^(1,8) Fixing the SL orientations provides a more precise description of the Cc position, indicating that the docking is reproducible.^(1,8) Note that the higher precision does not imply a result that is closer to the benchmark, in this case the position of Cc observed in the crystal structure, as is discussed below.

The distances between the Cc amide protons and the SL nitroxide radicals in the ensemble were compared to the experimental distances. Some discrepancies in the fit are to be expected since the data were obtained for the Cc-CcP complex in solution, which is known to consist of 30% encounter complex and 70% stereo-specific complex, while the structure comparison was done with the crystal structure, which consists only of the stereo-specific complex.⁽¹⁾ Nevertheless, there was a good overall fit between the back-calculated and experimental distances (Figure 5.8) and the fit was evaluated using the AV and Q-factors (Table 5.3).

TABLE 5.3 Q-factors and average violations (AV) for the fit of the back-calculated to the experimental distances, derived from inter-molecular PRE, between Cc amide protons and the oxygen in MTSL or pyMTSL at positions 38, 200 or 288 of CcP. The SL position was fit using a single conformer that was free to move (free) or fixed in the experimentally determined, most favourable orientation (fixed exp.).

Position		MTSL		pyMTSL	
		1 conformer free	1 conformer fixed exp.	1 conformer free	1 conformer fixed exp.
C38	AV (Å)	1.98	2.55	2.01	2.53
	Q-factor	0.10	0.15	0.10	0.14
C200	AV (Å)	2.13	2.51	1.88	2.19
	Q-factor	0.10	0.12	0.09	0.12
C288	AV (Å)	1.21	2.17	1.44	2.04
	Q-factor	0.07	0.12	0.08	0.11

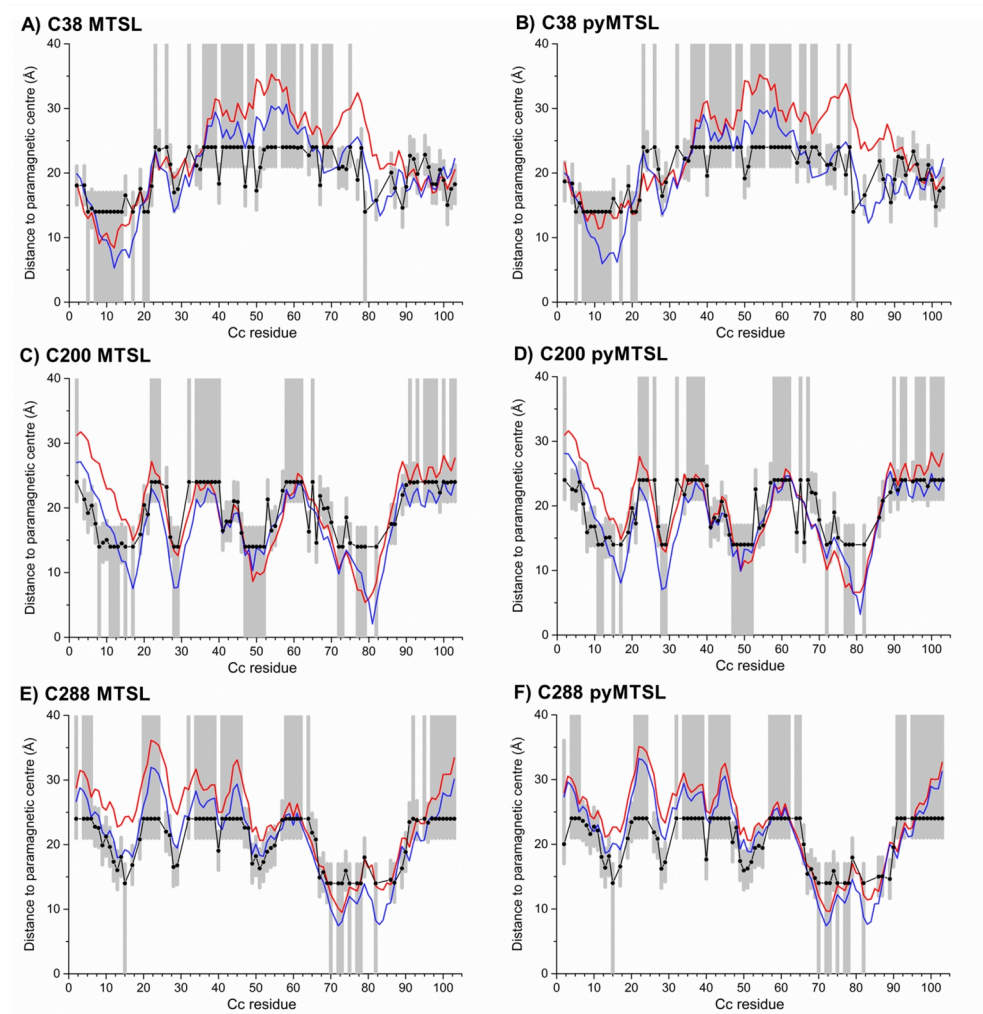


FIGURE 5.8 Experimental and back-calculated distances between Cc amide protons and the paramagnetic centre in MTSL or pyMTSL attached at positions 38 (**A,B**), 200 (**C,D**) or 288 (**E,F**) on CcP and plotted against the Cc residue number. The black circles and grey areas represent the experimental distances and error margins, respectively. The average back-calculated distances are shown as lines for a single fixed SL rotamer (red) or a single free SL rotamer (blue). The PRE were measured while CcP was in a 1:1 complex with Cc.

From the AV and Q-factors, it is clear that a freely rotating conformer much better describes the inter-molecular PRE data than does a single fixed conformer, despite the fixed position being determined experimentally with intra-molecular PRE data. This is not surprising since there are more degrees of freedom during the docking in this case; both the SL orientation and the position of Cc are fit to the inter-molecular PRE data when the SL is free to move.

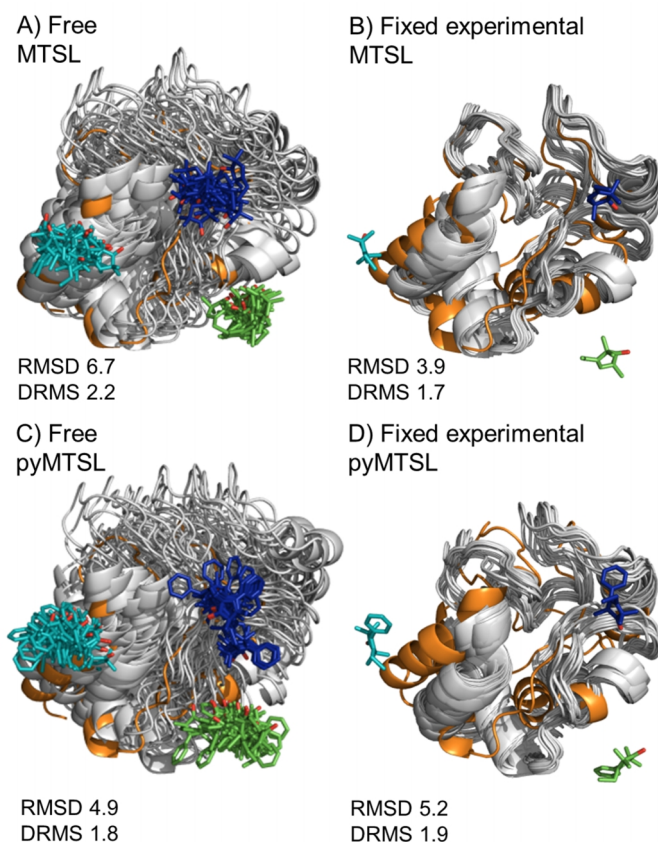


FIGURE 5.9 Comparison of the Cc positions as viewed from spin labelled CcP. The Cc position in the crystal structure (orange; PDB entry 2PCC)⁽¹⁶⁹⁾ is compared to the twenty lowest energy solutions for docking Cc to CcP (grey ribbons), driven by intermolecular PRE data using a single conformer that was free to move during DD (**A,C**) or that was fixed in the experimentally determined most favourable orientation (**B,D**). The SLs (MTSL **A, B**; pyMTSL **C, D**) are shown in sticks at positions 38 (teal), 200 (blue) and 288 (green), with the nitroxide oxygen atoms in red.

The results were also compared to the stereo-specific orientation found in the crystal structure (Figure 5.9). Overall, the Cc positions were similar to that of the crystal structure, although the calculated Cc ensemble was rotated slightly around the stereo-specific binding interface in all data sets. The RMSD of the C α atoms was calculated by first generating an average structure from all the Cc orientations, done by taking the linear average of the individual structures, and then comparing that to the stereo-specific orientation in the crystal structure. The RMSD is sensitive to differences caused by both rotation and translation, while the DRMS is mainly sensitive to translation. The DRMS is calculated by determining the C α -C α distance matrix of for all C α pairs from the two structures and then taking the root-mean-

square deviation (equation 5.2). The DRMS is always smaller than the RMSD and a large RMSD in combination with a small DRMS indicates that the two structures are mostly rotated relative to each other.⁽²⁹⁸⁾

For both fixed and free MTSL and pyMTSL, the RMSD values were much higher than the DRMS values, indicating that a large proportion of the difference between the calculated Cc orientations compared to the crystal structure is due to rotation of Cc, while the binding interface is similar. For MTSL, fixing the SLs in the experimentally determined most favourable orientations improved the fit to the crystal structure and decreased the DRMS from 2.2 Å for the freely rotating SLs to 1.7 Å with the SLs fixed (Figure 5.9, A & B). Also, the RMSD is significantly reduced from 6.6 Å to 3.9 Å. This indicates that for MTSL, experimentally determining the most favourable SL orientations prior to protein docking improved the accuracy of the final structure. For pyMTSL, the DMRS values were similar in both cases, with a DRMS of 1.8 Å and 1.9 Å when using freely rotating and fixed SLs, respectively.

SL CONFORMER ENSEMBLE

The protein docking calculations were repeated using the same inter-molecular PRE data and an ensemble of four SL conformers. Previous PRE studies on the Cc-CcP complex used an ensemble of four MTSL tags that were either fixed in selected positions during rigid body docking⁽¹⁾ or were free to move during a subsequent DD step.⁽¹³²⁾ Both approaches were repeated here along with a rigid body docking in which the SLs were fixed in the experimentally determined most favourable conformer orientations, using the orientations found in the lowest energy solution obtained with the intra-molecular PRE data. Cc was docked to CcP using inter-molecular PRE data and an ensemble was generated from the twenty lowest energy solutions for MTSL (Figure 5.10) and pyMTSL (Figure 5.11). As observed when using a single SL, allowing the SLs to rotate during docking resulted in a large, dispersed ensemble of the twenty lowest energy solutions for both MTSL and pyMTSL, while fixing the SL conformer orientations produced a well-defined position for Cc. The back-calculated distances between the Cc amide protons and the SL nitroxide radical oxygens were similar when using either free or fixed SL conformers and there was a good overall agreement with the experimental distances (Figure 5.12) and the fits were evaluated using the AV and Q-factors (Table 5.4).

TABLE 5.4 Q-factors and average violations (AV) for the fit of the back-calculated to the experimental distances, based on inter-molecular PRE, between CcP amide protons and the paramagnetic centre in MTSL or pyMTSL at positions 38, 200 or 288 when the SL positions were fit using ensembles of four conformers that were free to move (free), fixed in selected positions (fixed sele.) or fixed in the experimentally determined most favourable orientations (fixed exp.).

Position		MTSL			pyMTSL		
		4 free	4 fixed sele.	4 fixed exp.	4 free	4 fixed sele.	4 fixed exp.
C38	AV (Å)	2.35	2.53	2.66	2.00	2.18	2.40
	Q-factor	0.12	0.14	0.15	0.10	0.11	0.13
C200	AV (Å)	1.93	2.96	2.17	2.00	2.10	2.03
	Q-factor	0.10	0.14	0.12	0.08	0.11	0.11
C288	AV (Å)	1.32	1.51	2.24	1.60	1.80	2.32
	Q-factor	0.08	0.09	0.12	0.09	0.10	0.12

Again, the best fits were found when the SL orientations were free to move during DD; i.e., the SL conformer orientations were fit to the PRE data along with the position of Cc. This indicates that, although fixing the SL orientations results in a very precise determination of the Cc position, using four fixed positions for the nitroxide radical and, thereby, limiting the degrees of freedom, cannot describe the observed PRE data completely.

Surprisingly, optimizing the SL positions using experimental data did not yield a better fit to the inter-molecular PRE data than simply selecting four orientations from the sterically allowed possibilities (see Material and Methods for details of how the conformers were selected). The back-calculated distances are determined using the orientation of the complex found in the X-ray crystal structure, assuming that the orientation of the complex in solution is very similar to the one observed in the crystalline state. However, it is known that the complex is in fast exchange on the NMR timescale between this stereo-specific orientation, accounting for 70% of the complex in solution, and a more dynamic the encounter state, accounting for the remaining 30%.^(14, 18, 19) The experimental PREs are a (non-linear) average of all orientations in solution, so comparing them to only the stereo-specific orientation will limit the quality of the fit. Nevertheless, the overall fit to the PRE data was reasonably good for all conditions.

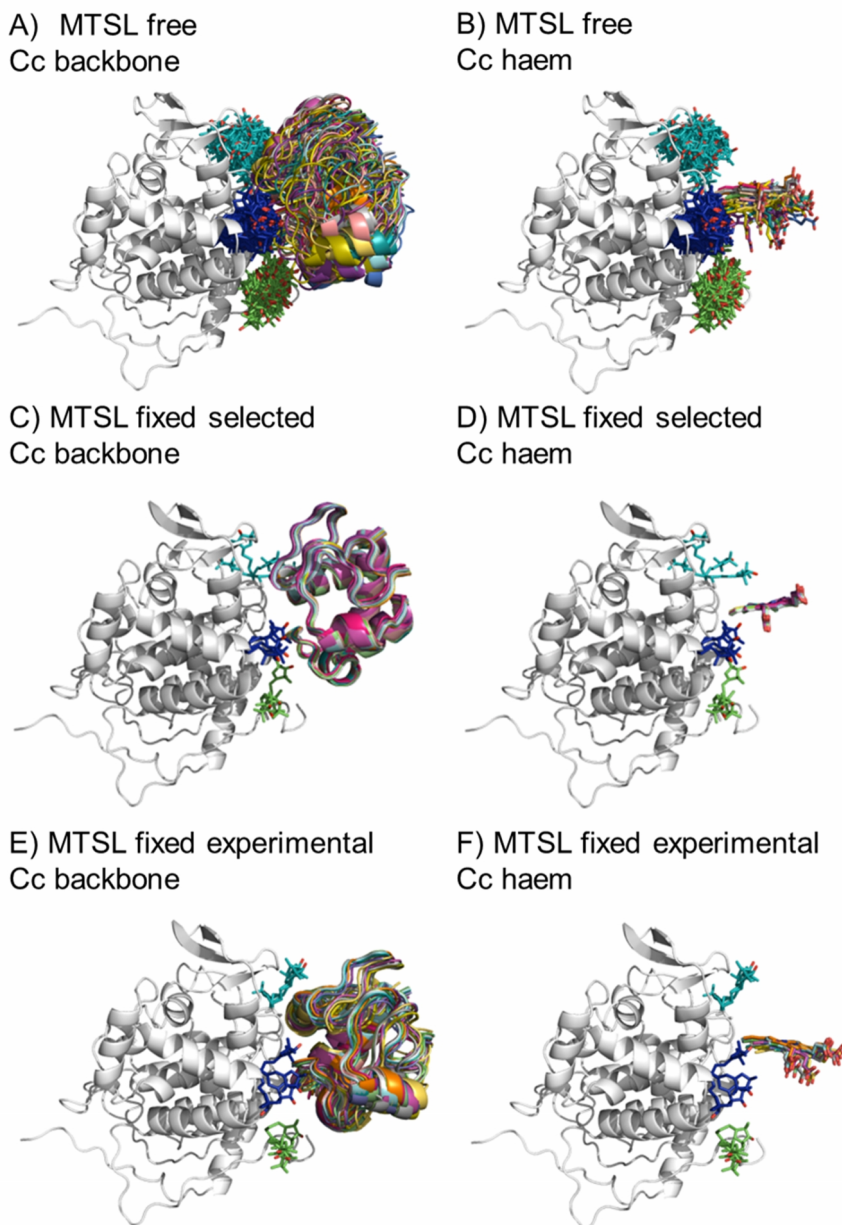


FIGURE 5.10 Twenty lowest energy solutions for docking Cc to CcP driven by intermolecular PRE data using an ensemble of four MTSL conformers that were free to move during DD (**A,B**) or that were fixed in selected positions (**C,D**) or experimentally determined most favourable orientations (**E,F**). CcP is shown in grey ribbons and Cc is shown in multi-coloured ribbons. The SLs are shown in sticks at positions 38 (teal), 200 (blue) and 288 (green), with the nitroxide oxygen atom in red and the Cc haem group is shown in multi-coloured sticks. The docking was done using structures from PDB entry 2PCC.⁽¹⁶⁹⁾

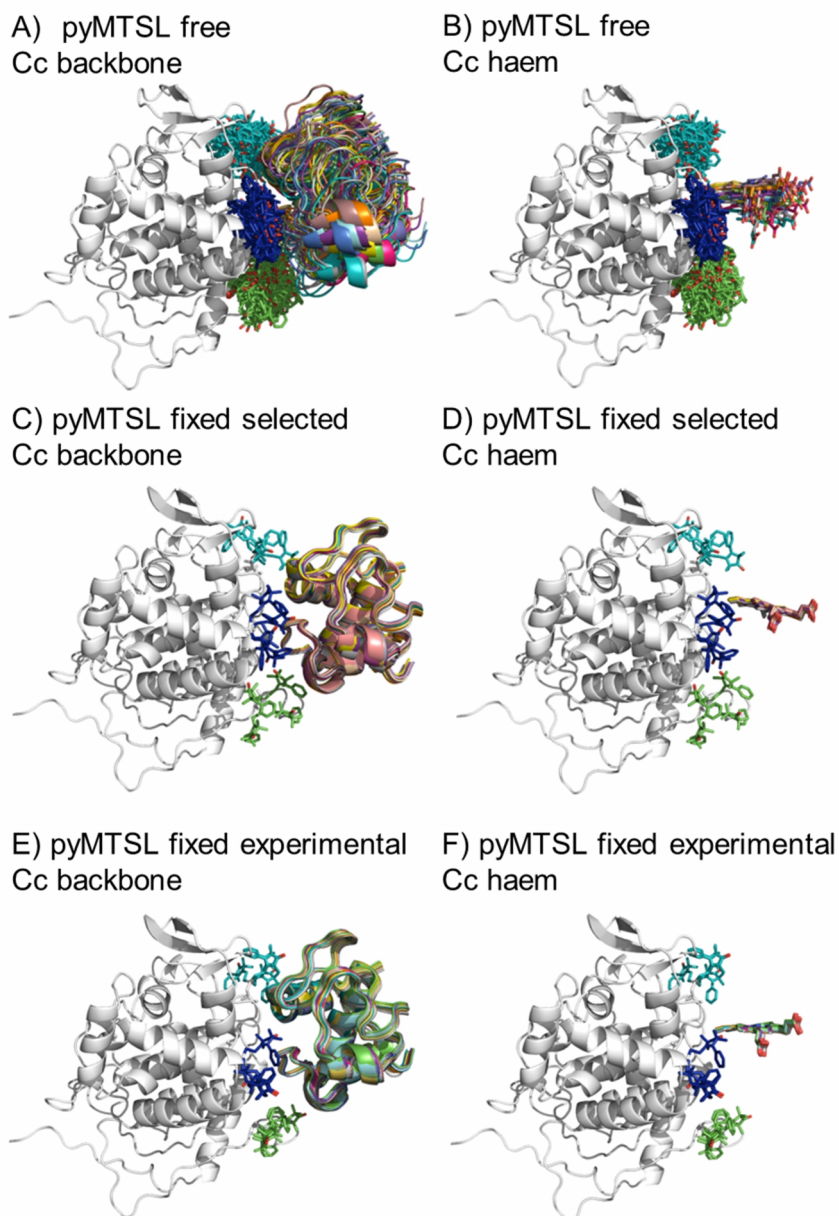


FIGURE 5.11 Twenty lowest energy solutions for docking Cc to CcP driven by intermolecular PRE data using an ensemble of four pyMTSL conformers that were free to move during DD (**A,B**) or that were fixed in selected positions (**C,D**) or experimentally determined most favourable orientations (**E,F**). CcP is shown in grey ribbons and Cc is shown in multi-coloured ribbons. The SLs are shown in sticks at positions 38 (teal), 200 (blue) and 288 (green), with the nitroxide oxygen atom in red and the Cc haem group is shown in multi-coloured sticks. The docking was done using structures from PDB entry 2PCC.⁽¹⁶⁹⁾

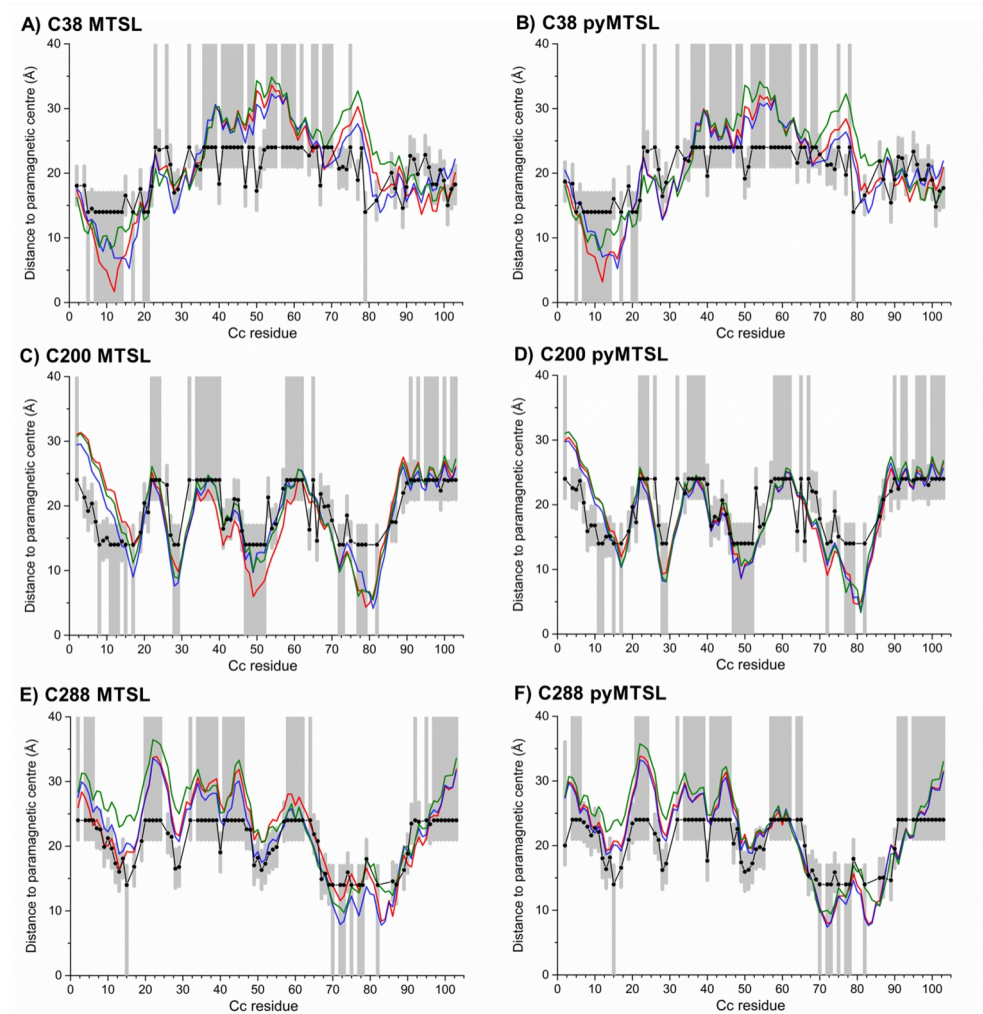


FIGURE 5.12 Experimental and back-calculated distances between Cc amide protons and the paramagnetic centre in MTSL or pyMTSL attached at positions 38 (**A,B**), 200 (**C,D**) or 288 (**E,F**) on CcP C128A plotted against the Cc residue number. The black circles and grey areas represent the experimental distances and error margins, respectively. The average back-calculated distances are shown as lines for an ensemble of four SL rotamers that were free to rotate (blue), fixed in random positions (red) or fixed in experimentally determined most favourable orientations (green). The PRE were measured while CcP was in a 1:1 complex with Cc.

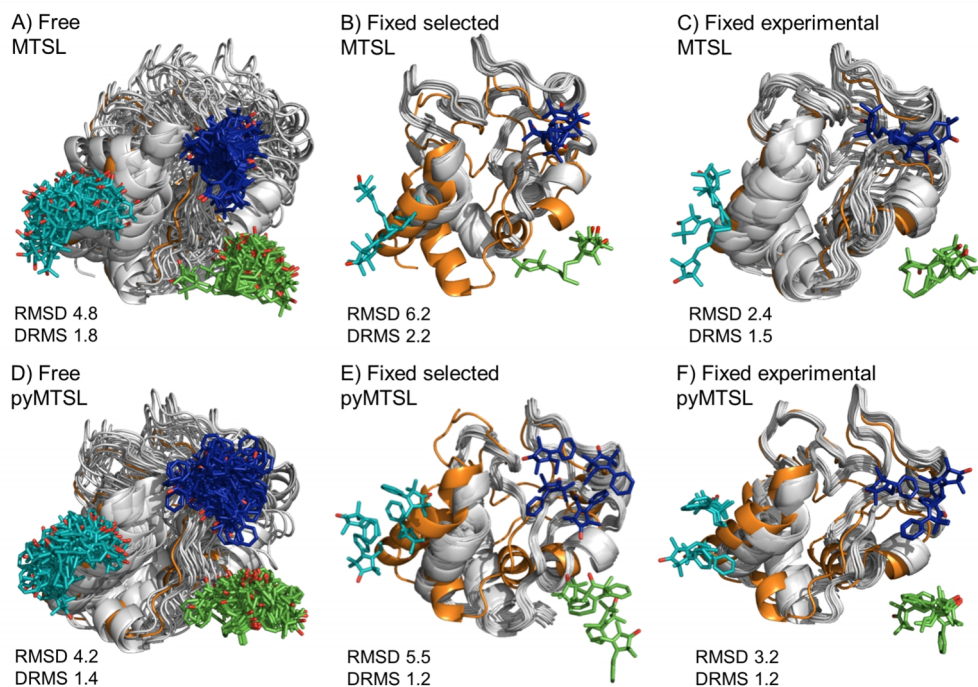


FIGURE 5.13 Comparison of the Cc positions as viewed from spin labelled CcP. The Cc position in the crystal structure (orange; PDB entry 2PCC)⁽¹⁶⁹⁾ is compared to the twenty lowest energy solutions for docking Cc to CcP (grey ribbons), driven by intermolecular PRE data using an ensemble of four conformers that were free to move during DD (**A,B**) or that were fixed in selected positions (**C,D**) or experimentally determined most favourable orientations (**E,F**). The SLs are shown in sticks at positions 38 (teal), 200 (blue) and 288 (green), with the nitroxide oxygen atoms in red and the Cc haem group shown in multi-coloured sticks.

The results for docking Cc to CcP were also compared to the crystal structure of the stereo-specific complex (Figure 5.13). As shown previously,^(287, 288) the best results were obtained for MTSL when using multiple conformers. Furthermore, as seen when using a single SL conformer, many of the resulting Cc orientations are rotated with respect to the stereo-specific binding site resulting in high RMSD values but small DRMS values for the ensembles. For MTSL, fixing the conformers in the selected orientations produced the worst fit the with crystal structure (RMSD=6.7 Å; DRMS=2.2 Å). The fit improved when the SLs could freely rotate (RMSD=4.8 Å; DRMS=1.8 Å) but the best fit was obtained when using the experimentally determined conformer orientations (RMSD=2.5 Å; DRMS=1.5 Å). This indicated that the most accurate description of the protein complex was achieved by predetermining the conformer orientations using intra-molecular PRE, even though the quality of fit to the experimental data was not the best for this solution.

In a previous study, PRE from MTSL at positions 38, 200 and 288 were fixed in selected positions while Cc was docked to CcP.⁽¹⁾ The same conformers were used in this study. An RMSD with the crystal structure of 2.2 Å was found in that study, which is much smaller than the RMSD values that we observed, of 6.7 Å. This large discrepancy is likely due to small differences in the experimental PRE and the fact that normalization of the $I_{\text{para}}/I_{\text{dia}}$ data (as described in the materials and methods) was not done in the previous study. Although the $I_{\text{para}}/I_{\text{dia}}$ ratio for residues unaffected by PRE is expected to be 1.0, we frequently observe values that are on average slightly higher or slightly lower. The reason for this is unclear but is likely due to slight differences between the diamagnetic and paramagnetic samples. The transverse relaxation rate is very sensitive to the exact fraction bound, because the rotational correlation time of the CcP-Cc complex is much larger than that of free Cc. If the concentrations of the proteins vary slightly, for example, due to slight aggregation, the fraction bound Cc can differ between the paramagnetic and diamagnetic samples. Since it is a global effect on all residues, and the deviation in the average $I_{\text{para}}/I_{\text{dia}}$ ratio can be both larger and smaller than 1.0, non-specific contact between Cc and the paramagnetic protein cannot be the cause of this effect. This normalization was not done in the previous study⁽¹⁾ resulting in shorter distance restraints than were obtained in this study, particularly for position C38, which allowed Cc to find a position much closer to the SLs than was allowed in this study. These findings demonstrate that the outcome of the docking is very sensitive to relatively small differences in the experimental data set.

For pyMTSL, the DRMS values showed that using pyMTSL resulted in a very good fit to the stereo-specific state but the RMSD values were quite high. Therefore, the differences between the final Cc positions found after docking and the stereo-specific state were mainly due to rotation and not translation of Cc. When the SL was free to rotate, the resulting fit had an RMSD of 4.2 Å and DRMS 1.4 Å. Fixing the conformers in the selected orientations resulted in greater rotation of Cc relative to the stereo-specific state (RMSD=5.5 Å; DRMS=1.2 Å). Fixing the SLs in the experimentally determined orientations resulted in a lower RMSD of 3.2 Å and DRMS 1.2 Å, in line with the results for MTSL. Therefore, in this case, it was necessary to represent pyMTSL using multiple conformers to get the best results.

As was shown previously,⁽¹⁾ fixing the SL orientations during docking produces a highly defined ensemble of solutions. However, precision is of no importance in the absence of accuracy, which can be difficult to achieve when the fixed SL orientations must be selected without experimental data. The use of intra-molecular PRE to predetermine the most favourable SL orientations allowed improvement of the accuracy of structure determination

despite the fact that this did not necessarily improve the fit between the experimental and back-calculated data. In this case, we used the stereo-specific orientation of Cc observed in the crystal structure as the benchmark for accuracy under the assumption that this is also the main state of the complex in solution. However, as mentioned above, this stereo-specific state only accounts for 70% of the complex in solution, with the remaining 30% in the encounter state,^(1, 8, 35, 140, 204) resulting in discrepancies in the fit. Furthermore, under most conditions, significant rotation of Cc around the stereo-specific binding interface was observed. This is likely due to the fact that PRE are highly sensitive to minor states⁽⁵⁴⁾ giving them a disproportionately large influence on the final orientation of Cc determined by the docking calculations. Although fixing the SLs in the experimentally determined most favourable orientations helped to reduce this effect, we conclude that PRE do not yield very reliable restraints for determining protein orientation within a complex.

Much better results can be obtained when combining PCS or RDC with the PRE data.^(14, 31, 80, 104-108, 299, 300) This was demonstrated recently by Hiruma *et al.* for the cytochrome P450cam–putidaredoxin complex. Paramagnetic tags were placed at two locations on cytochrome P450cam as well as on one location on putidaredoxin and intermolecular PCS, RDC and PRE were obtained. The PCS and RDC back-calculated from the final, well-defined structure matched the experimental data very well, as did the PRE data for the one of the tag positions on cytochrome P450cam. For the second position, however, the experimental PRE were stronger than expected for many residues and the authors concluded that this was due to a minor state influencing the PRE results. Nevertheless, the use of PCS and RDC in combination with PRE data allowed for the structure of the major state to be successfully resolved, as judged by the subsequently determined crystal structure of the complex, which showed a 1.7 Å RMSD with the mean of the NMR ensemble of structures.⁽³⁰¹⁾ This work was done using different paramagnetic Ln³⁺ ions, which can be attached to the protein via double armed caged lanthanoid NMR probes (CLaNP), having the additional benefit of being highly immobilized and therefore limiting spatial averaging of the observed paramagnetic effects.^(85, 102)

For structural modelling paramagnetic restraints can also be combined with other NMR data, such as RDCs obtained using external alignment media and nuclear Overhauser effects (NOE).^(80, 104-108, 299) This was done recently by Shi *et al.* who used PRE data to complement both NOE and RDC data for the integral-membrane protein phospho-lamban (PLN). PLN is a small protein consisting of two helical domains linked by a flexible loop. The NMR structure of the monomeric protein in dodecylphosphocholine micelles had previously

been solved by their group using NOE and solvent PRE but the resulting ensemble of conformers in the membrane was very poorly defined.⁽³⁰²⁾ Combining the NOE data with RDCs resulted in 100 low energy structures that were highly refined (backbone RMSD=1.6 Å) but were grouped into four families exhibiting a four-fold degeneracy in the relative orientations of the two helices. By also incorporating PRE data into the structure refinement, the degeneracy of the RDC data was overcome and the correct family of structures was obtained with high resolution (backbone RMSD=1.2 Å).⁽³⁰⁰⁾ This highlights the benefit of combining multiple data sets when resolving the structure of dynamic and/or multi-domain proteins, especially when together they can provide both distance and orientation information.

This work has combined both intra- and inter-molecular PRE data to investigate the role of SL mobility on complex structure determination. While little difference was found between MTSL and pyMTSL, the accuracy of final results, as judged by similarity to the crystal structure of the complex, was highly dependent on the number and choice of SL conformer orientations used during the docking. It was also found that fixing the SL orientations during docking resulted in highly precise ensembles for the Cc position but that this level of precision was not correlated with a better match to the stereo-specific orientation of Cc. Although pre-determination of the favourable SL orientations using intra-molecular PRE data did help to improve the accuracy of docking results for the Cc-CcP complex, this did not necessarily improve the fit between the experimental and back-calculated data, so additional cases should be studied to assess the value of this technique for highly dynamic complexes. Overall, it seems that PRE determined distance restraints used in isolation are not ideal for determining the protein orientation within a dynamic complex and much better results can be obtained when combining PCS or RDC with the PRE data.⁽³⁰¹⁾

CHAPTER 6

Ultra-weak self-association between cytochrome *c* peroxidase molecules observed by paramagnetic NMR

ABSTRACT

Many biologically relevant interactions are weak as a consequence of intracellular localization. Binding partners will only meet when they are co-localized in the same compartment but local concentrations can be very high enabling the formation of weak complexes. With the development of PRE NMR it has become possible to characterize weak complexes experimentally and even detect complexes with affinities in the 1-25 mM range. Here we present evidence for an ultra-weak self-association between CcP molecules. In order to generate intramolecular PRE, nitroxide based SLs were attached at three positions around the stereo-specific binding interface on CcP. Multiple additional intermolecular PRE were also observed and these could be diminished by the addition of Cc. The interaction appears to occur between the stereo-specific binding interface on the “front” of CcP and several small positively charged patches on the “back” of CcP centred on residue K278. Since CcP has until now been considered as a monomer in solution, the existence of this ultra-weak self-association may have implications for experimental results, particularly when studying weak effects. This finding is also in line with growing evidence that many proteins exhibit a tendency for ultra-weak homo- or hetero- oligomerization interactions.

INTRODUCTION

Decades of research on protein-protein interactions has revealed valuable information about the structure and function of many protein complexes. However, the majority of this work has focused on proteins that form strong and often highly specific interactions while the importance of weakly associated complexes, such as the encounter complex, is becoming increasingly clear. These transient complexes are commonly used by nature to counter-balance the biological need for a specific interaction with the necessity of maintaining a high turnover rate.⁽⁵⁷⁾ This also extends to ultra-weak interactions, defined as having a K_D in the millimolar range,⁽³⁰³⁻³⁰⁵⁾ that are known to drive self-assembly of higher order homogeneous architectures like crystals, viral capsids and amyloid fibrils.⁽³⁰⁶⁻³⁰⁸⁾ They also play an important role in an array of cellular processes including rapid assembly/disassembly, protein maturation, reversible cell adhesion and cell signalling.⁽³⁰³⁾ While strongly associated protein complexes often consist of low energy, specific states that are easily isolated and studied, weakly associated protein complexes occupy several high-energy conformations over time.⁽³⁰⁹⁾ These conformations are lowly populated, transient and cannot be isolated making them practically invisible to conventional structural biology techniques.⁽⁵⁹⁾

The development of new pNMR techniques has enabled characterization of these transient states in populations as low as 0.5%.⁽⁶²⁾ PRE is particularly well-suited to studying lowly populated states as the dipolar interaction with the unpaired electron is very strong and the strength of the PRE falls off with the sixth power of the distance between the paramagnetic centre and observed nucleus, making PRE extremely distance dependent.^(1, 28, 29) PRE has been successfully used to characterize several protein complexes/encounter complexes^(1, 14, 31, 34, 54, 301, 310) including the Cc-CcP encounter complex.^(1, 8, 9, 140, 204) It has also been used to study protein-DNA complexes^(29, 289, 311) as well as large scale domain motions^(288, 312) and transient structures in unfolded and intrinsically disordered proteins.^(105, 313-316)

In 2008, PRE was used for the first time to visualize ultra-weak self-association ($K_D \geq 15$ mM) between HPr molecules that could not be observed by other techniques. Paramagnetic EDTA-Mn²⁺ tags were placed at three positions on the surface of unlabelled HPr. This protein was mixed 1:1 (300 μ M each) with [¹⁵N]-labelled HPr and PRE were measured. An ultra-weak self-association was observed, representing a population of 1%, which disappeared when the physiological HPr binding partner, enzyme I, was added. Furthermore, this interaction could be modulated by changes in the ionic strength or with charge mutations.⁽³⁰⁴⁾ In the same year, PRE was used to show an ultra-weak interaction between the N-terminal extension of the

HIV-1 protease precursor and the protein's active site resulting in autocleavage and maturation of the protein. The ultra-weak encounter complex ($K_D = 3\text{-}6$ mM) was found to represent 3-5% of the total population in a concentration of 200 μM .⁽³⁰⁵⁾ This technique has since been applied to several ultra-weak complexes⁽³¹⁷⁻³¹⁹⁾ with K_D values as high as 25 mM.⁽³²⁰⁾ Together, these studies demonstrated the use of PRE for the observing weak interactions of self-associations that are difficult to visualize with conventional methods.

In our previous work using nitroxide based SLs to obtain intramolecular PRE data on CcP (Chapter 5), we observed multiple unexpected PRE effects for residues further than 24 Å from the SL, the PRE limit for nitroxide radicals.⁽⁶²⁾ Here we show that these additional PRE are in fact intermolecular effects generated by an ultra-weak self-association between the CcP molecules. These effects are diminished upon the addition of Cc. Until now, CcP had been assumed to exist as a monomer in solution; the existence of this ultra-weak CcP-CcP complex may have implications for scientific results and will need to be taken into account when designing future experiments.

MATERIALS & METHODS

PROTEIN SAMPLE PREPARATION

The yeast CcP C128A and the 38C, 200C and 288C mutants^(1, 321) and yeast iso-1-Cc^(124, 125) were expressed and purified as published previously (Chapters 3). MTS and MTSL tags were obtained from Toronto Research Chemicals (North York, ON, Canada) and pyMTSL was synthesized according to the published protocol by Dr. Wei-min Liu.⁽²⁹¹⁾ The SLs were stored as 100 mM stocks dissolved in DMSO at 4 °C prior to use. The CcP mutants were tagged with MTS, MTSL or pyMTSL, as described previously (Chapter 4).^(1, 140) The tagging efficiency was determined by mass spectroscopy to be essentially 100%.

NMR SPECTROSCOPY

All NMR samples contained 20 mM NaPi, 100 mM NaCl, 6% D₂O, pH 6.0. The pH of the samples was adjusted to 6.00 ± 0.05 , with small aliquots of 0.5 M HCl or NaOH. To determine the optimal CcP concentration, 2D [¹⁵N,¹H] TROSY-HSQC^(224, 237-241) spectra were obtained with 1024 and 128 complex points in the direct and indirect dimensions, respectively, for 400-800 μM double labelled [¹⁵N, ¹³C] CcP at 293 K. Measurements were performed at ¹H Larmor frequencies of 600 MHz on a Bruker Avance III spectrometer equipped with a TCI-Z-

GRAD CryoProbe (Bruker, Karlsruhe, Germany). The data were processed using Topspin 3.1 (Bruker, Karlsruhe, Germany).

For intra-molecular PRE measurements, NMR samples contained 300 μM unlabelled Cc WT and 300 μM double labelled [^{15}N , ^2H] NC38 or TC288 or ^{15}N -labelled NC200 CcP with either MTS, MTSL or pyMTSL tags attached in 20 mM NaPi, 100 mM NaCl, 6% D_2O , pH 6.0. 2D BEST-TROSY-HSQC experiments⁽²⁷⁹⁾ were recorded on a Bruker AVIII HD spectrometer equipped with a $^1\text{H}[^{13}\text{C}/^{15}\text{N}]$ TCI-cryoprobe operating at a Larmor frequency of 850 MHz at 293 K with 1024 and 100 complex points in the ^1H and ^{15}N dimensions, respectively. The data were processed using Topspin 3.2 (Bruker, Karlsruhe, Germany). All NMR data were analyzed using CCPN Analysis 2.1.5.⁽²⁴²⁾

PRE ANALYSIS

The intensity ratio of the amide resonances in the spectra of the paramagnetic (MTSL) and diamagnetic (MTS) samples ($I_{\text{para}}/I_{\text{dia}}$) was measured and normalized as described previously.⁽⁸⁾ The paramagnetic contribution to the transverse relaxation rate, $R_{2,\text{para}}$, was calculated as described previously (Chapter 4).^(8, 77, 140) For the amide peaks that disappeared in the paramagnetic spectrum, an upper limit for I_{para} was set to two standard deviations of the noise level of the spectrum.⁽¹⁴⁰⁾ The calculated $R_{2,\text{para}}$ values were then converted into distances as described previously,⁽⁸⁾ see equation 4.1, with a fraction bound of 1.0 for intramolecular PRE on CcP. The value of τ_c for the complex was previously estimated to be 16 ns.^(1, 8) The residue depth from the protein surface was calculated using the DEPTH server^(322, 323) with 25 solvating cycles, 4.2 Å solvent neighbourhood radius and a minimum of 2 neighbourhood waters.

RESULTS & DISCUSSION

To enable the backbone amide resonance assignment of CcP (Chapter 3),⁽¹⁴⁰⁾ we determined the optimal CcP concentration for NMR samples. Previous NMR studies on the complex between CcP and Cc were done using a 1:1 ratio at 300 μM of each protein.^(1, 8, 35) Cc is remarkably soluble (~ 100 mg/mL)⁽¹²⁰⁾ so the NMR sample concentration was previously limited by the solubility of CcP. However, altering the purification protocol (Chapter 3)⁽¹⁴⁰⁾ had resulted in protein that was stable at much higher concentrations. It was hoped that indeed higher concentrations could be used for the NMR spectroscopy to take advantage of the

increased signal intensity that would provide. To determine the optimal CcP concentration for NMR, 2D [^{15}N , ^1H] TROSY-HSQC spectra were obtained for samples containing 400-800 μM CcP (Figure 6.1).

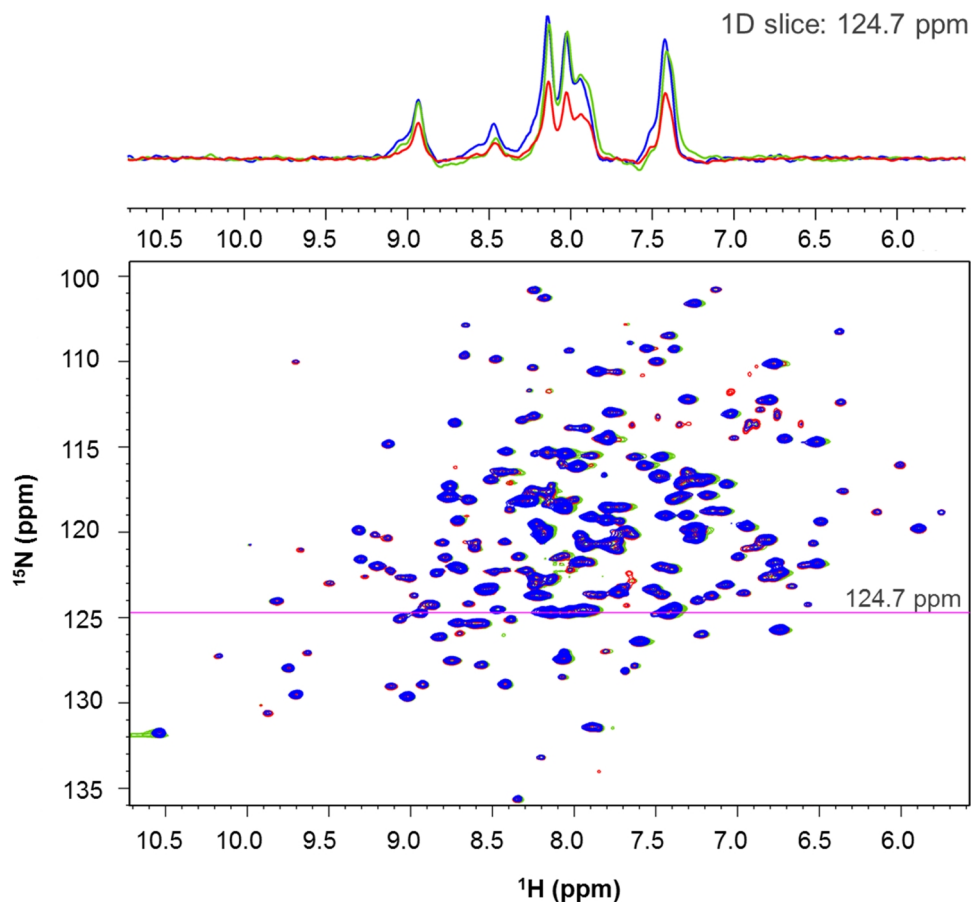


FIGURE 6.1 2D [^{15}N , ^1H] TROSY-HSQC spectra of 400 μM (blue), 600 μM (green) or 800 μM (red) double labelled [^{13}C , ^{15}N] CcP C128A with 1D overlay (above) showing relative ^1H peak intensities at 124.7 ppm in the ^{15}N dimension (pink line). Experiments were performed at 20 $^\circ\text{C}$ in 20 mM NaPi, 100 mM NaCl, pH 6.0.

In a sample that does not aggregate, the signal-to-noise ratios should be proportional to the sample concentration. However, although no visible aggregation was observed in the sample, the 1D traces of the spectra showed little or no peak intensity increase when going from 400 μM to 600 μM and a large decrease in peak intensity when the sample concentration was further increased to 800 μM (Figure 6.1). This suggests that indeed aggregation of CcP was occurring in the sample. When proteins aggregate, the intensity increase at higher concentrations is counteracted by enhanced nuclear relaxation, resulting in intensity loss due to the larger rotational correlation time of the aggregate as compared to the monomeric state of the protein. Weak self-association has been seen in other proteins at concentrations 200-500 μM .^(304, 305, 318, 319)

Further evidence for weak CcP-CcP interactions was observed during a study on the use of intra- and intermolecular PRE in combination to improved protein docking results (Chapter 5).⁽³²¹⁾ We attached MTSL SLs containing a paramagnetic nitroxide radical at three positions on CcP around the stereo-specific binding interface for Cc. Nitroxide based SLs generate measurable PRE effects up to 24 Å for a protein the size of CcP.⁽⁶²⁾ Thus, no intramolecular PRE were expected beyond this limit. When we measured the intramolecular PRE, we found, however, many unexpected effects for residues spread across the CcP sequence. By converting the PRE into the distances between the nitroxide oxygen atom and the amide hydrogens of CcP, the intra- and intermolecular PRE could be distinguished (Figure 6.2). In Figure 6.2, the theoretical distances, back predicted from the experimentally determined most favourable SL orientations of a four SL ensemble, are plotted in a dark blue line, which represents where the intramolecular PRE were expected. The blue dots represent residues within 24 Å of the nitroxide oxygen atom, for which mainly intramolecular PRE are expected. These PRE can represent a mixture of intra- and intermolecular PRE but the intramolecular PRE is expected to dominate. For these residues there is a very nice fit between the experimental and back-predicted intramolecular PRE. Since the remaining nuclei are further than 24 Å away from the nitroxide oxygen atom, no intramolecular PRE were expected and thus the red dots likely represent intermolecular PRE generated by weak CcP-CcP interactions.

For intermolecular PRE, the effects should be mainly on the surface of the protein while intramolecular PRE will also occur for residues buried within the protein. The intra- and intermolecular PRE derived distances between the nitroxide oxygen atom in the SL and the CcP amide hydrogens was compared to the distance of the residues from the protein surface (Figure 6.3).

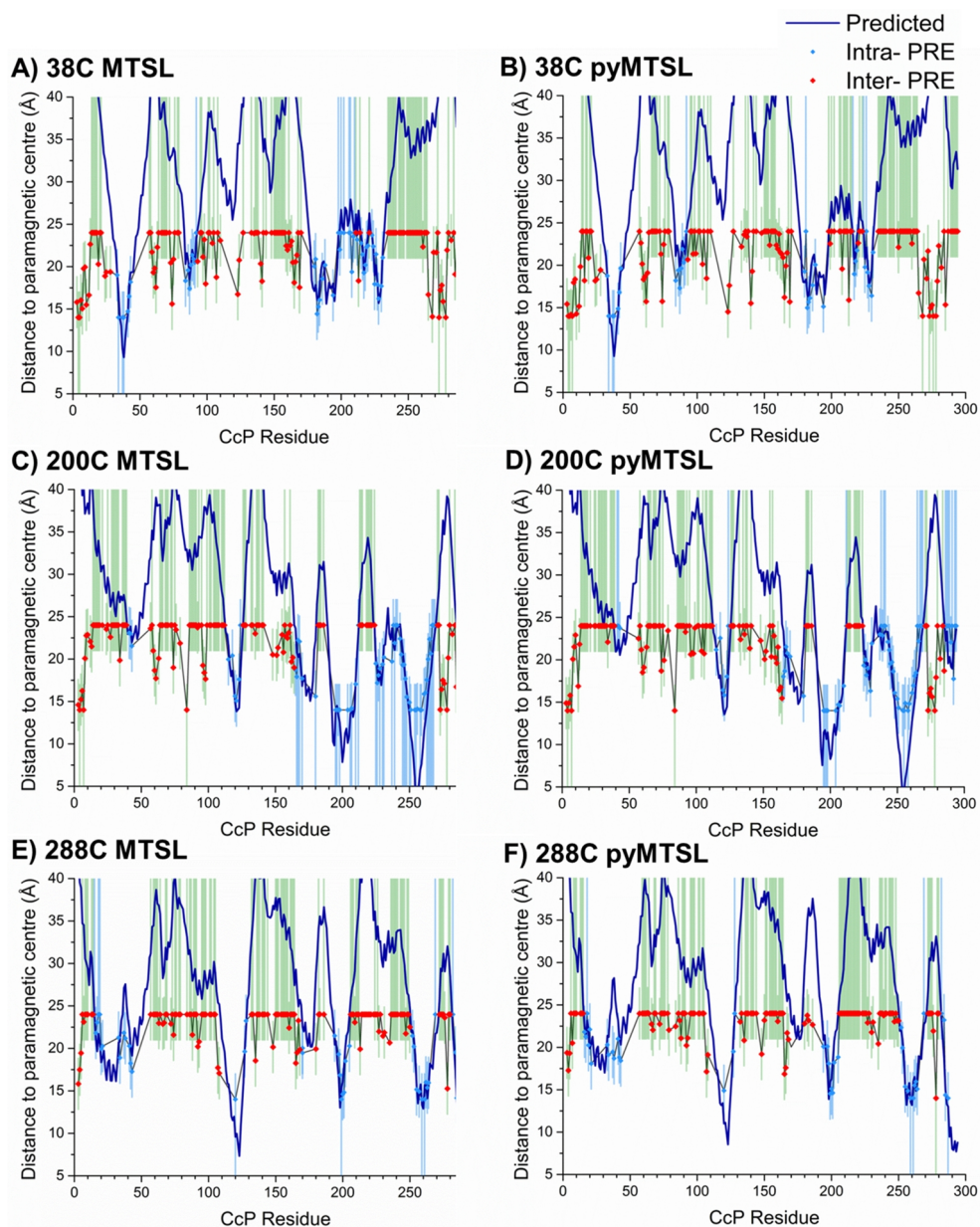


FIGURE 6.2 Inter- (red) and intra- (light blue) PRE derived distances observed for CcP amide hydrogen atoms generated by a nitroxide radical in (py)MTSL attached at positions 38C (**A,B**), 200C (**C,D**) or 288C (**E,F**). The theoretical distances, back predicted from the experimentally determined most favourable SL orientations of a four SL ensemble,⁽³²¹⁾ are plotted as a dark blue line and the errors are shown in green and blue bars for the inter- and intra- PRE, respectively.

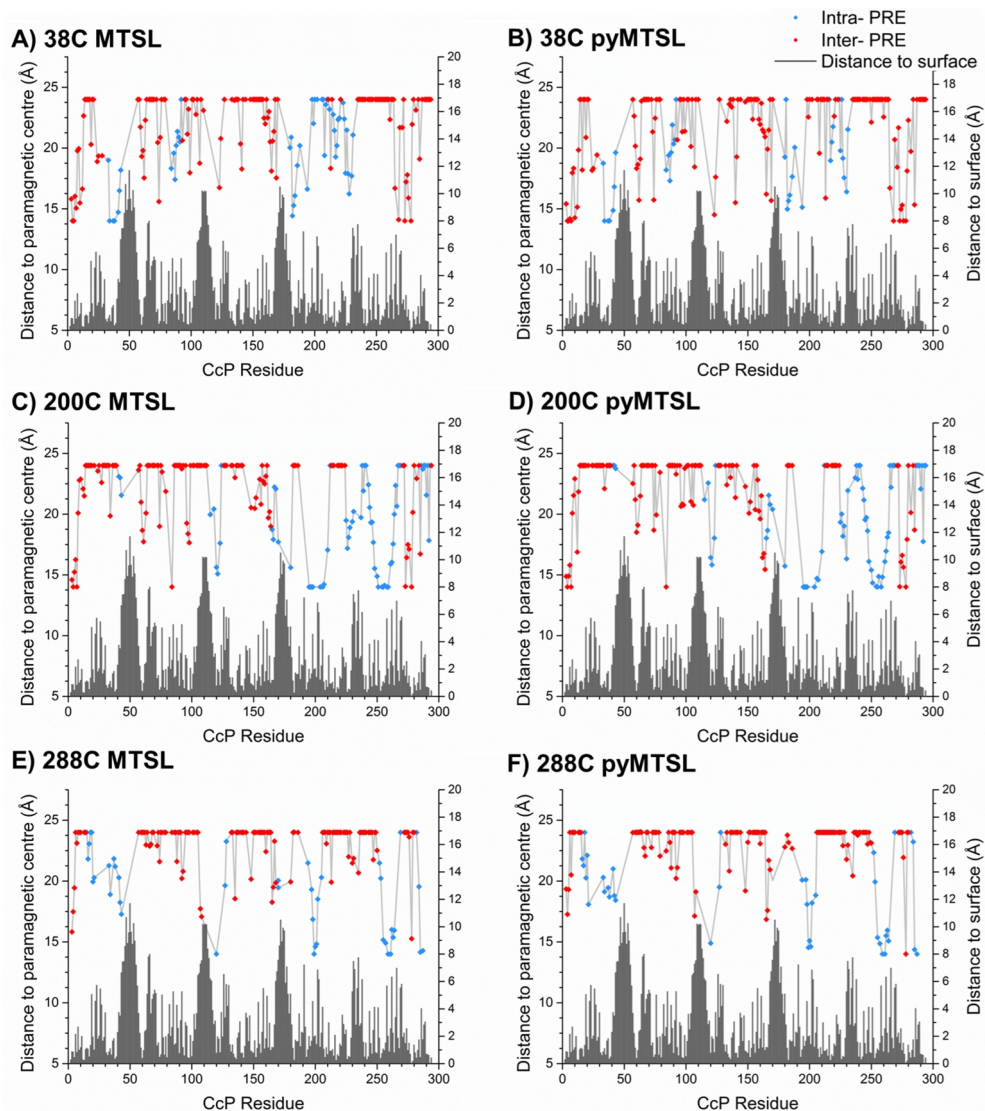


FIGURE 6.3 Inter- (red) and intra- (light blue) PRE derived distances between CcP amide hydrogen atoms generated and the nitroxide radical in (py)MTSL attached at positions 38C (**A,B**), 200C (**C,D**) or 288C (**E,F**) compared to the distance of the residues from the surface of CcP (grey bars). The residue depth from the protein surface was calculated using the DEPTH server.^(322, 323)

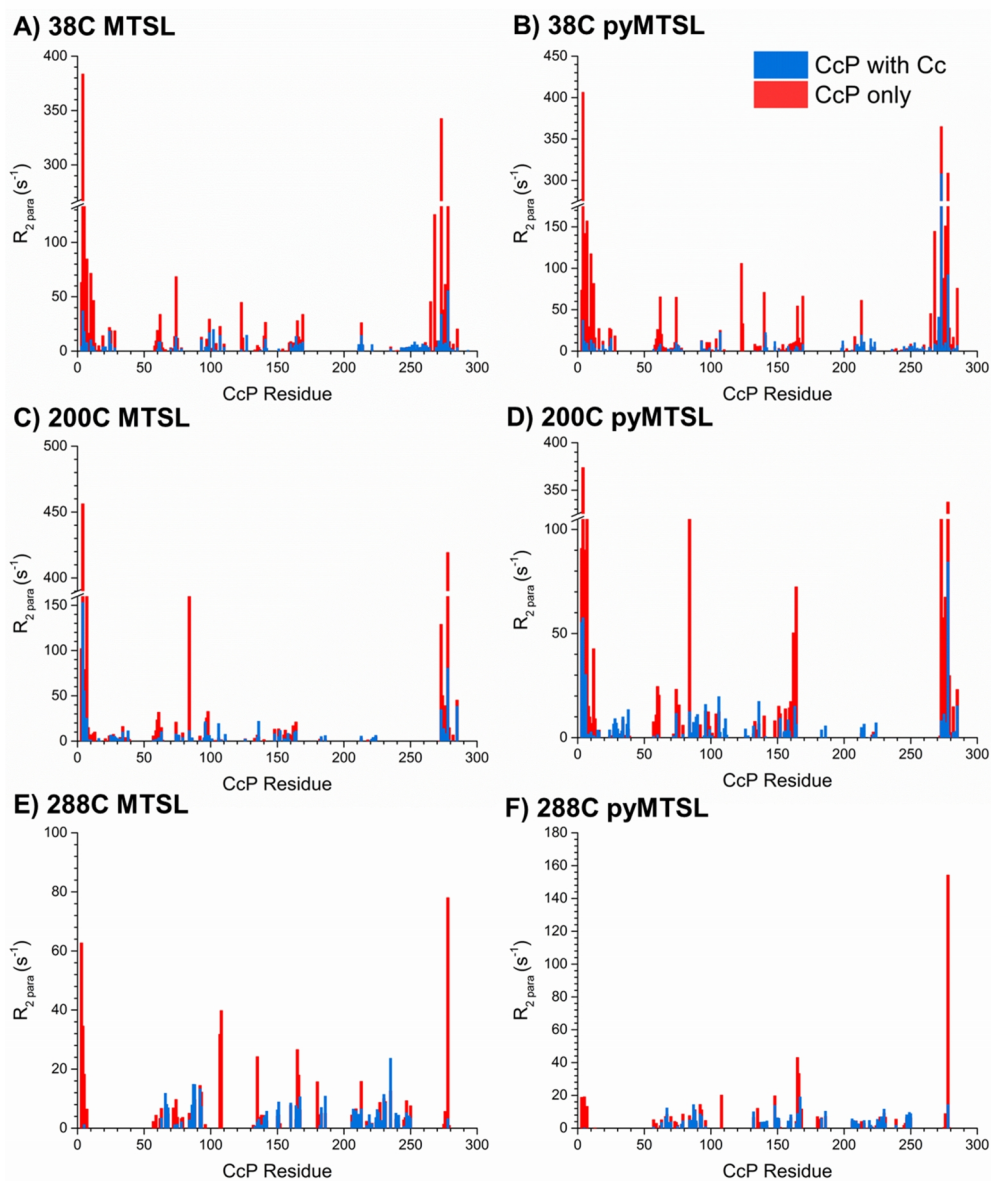


FIGURE 6.4 Comparison of $R_{2\text{para}}$ for CcP amide nitrogen atoms in a sample containing only CcP (red) or in a 1:1 with Cc (blue) generated by a nitroxide radical in (py)MTSL attached to CcP at positions 38C (**A,B**), 200C (**C,D**) or 288C (**E,F**). $R_{2\text{para}}$ values $> 128 \text{ s}^{-1}$ cannot be measured accurately and are estimates.

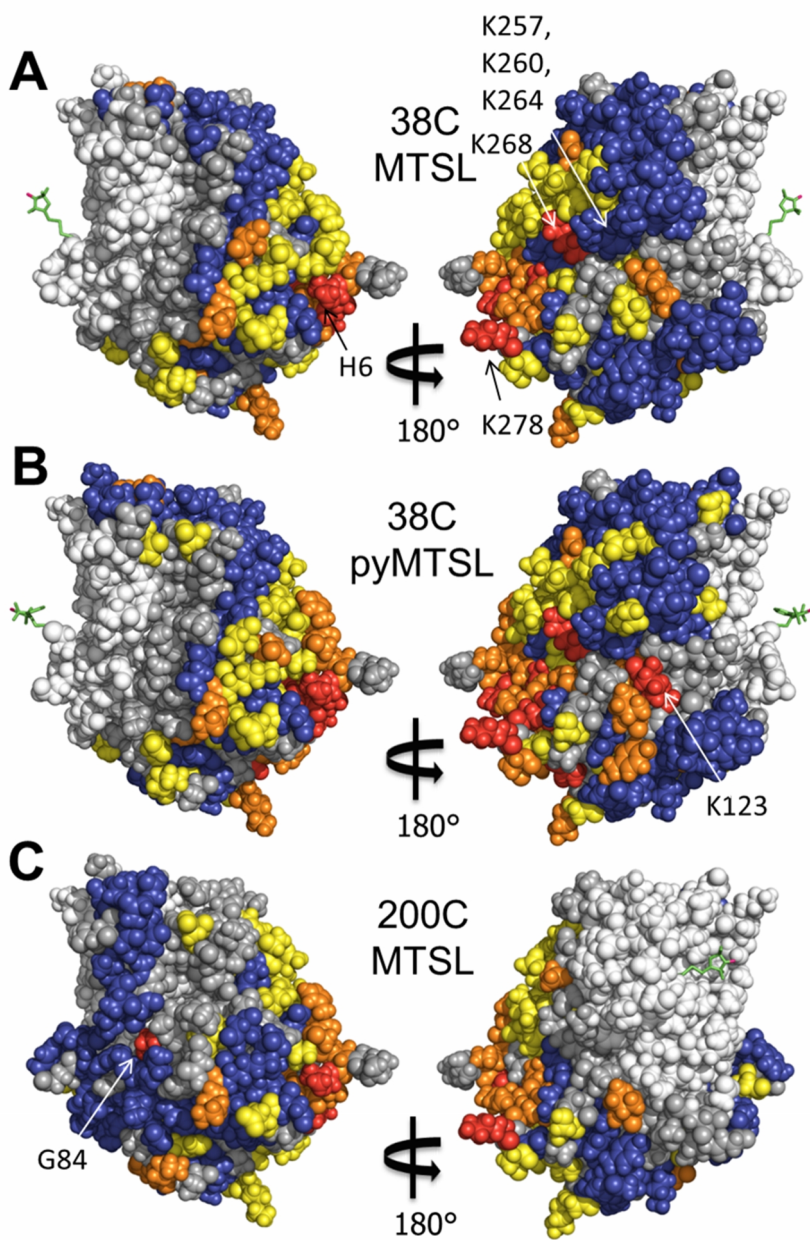


FIGURE 6.5 Intermolecular PRE map for CcP with MTSL or pyMTSL at positions 38C (**A,B**), 200C (**C,D**) or 288C (**E,F**). The PRE effects are colour-coded on CcP (PDB-entry 2PCC)⁽¹⁶⁹⁾ and the (py)MTSL is shown in green sticks with the oxygen atom in red. Intramolecular PRE are expected within 24 Å of the nitroxide oxygen atom so these residues are coloured white. Residues with $\Gamma_{2, \text{para}} \geq 100 \text{ s}^{-1}$ are red, $20 \text{ s}^{-1} < \Gamma_{2, \text{para}} < 100 \text{ s}^{-1}$ are orange, $5 \text{ s}^{-1} < \Gamma_{2, \text{para}} < 20 \text{ s}^{-1}$ are yellow, $\Gamma_{2, \text{para}} \leq 5 \text{ s}^{-1}$ are blue and with no data are grey.

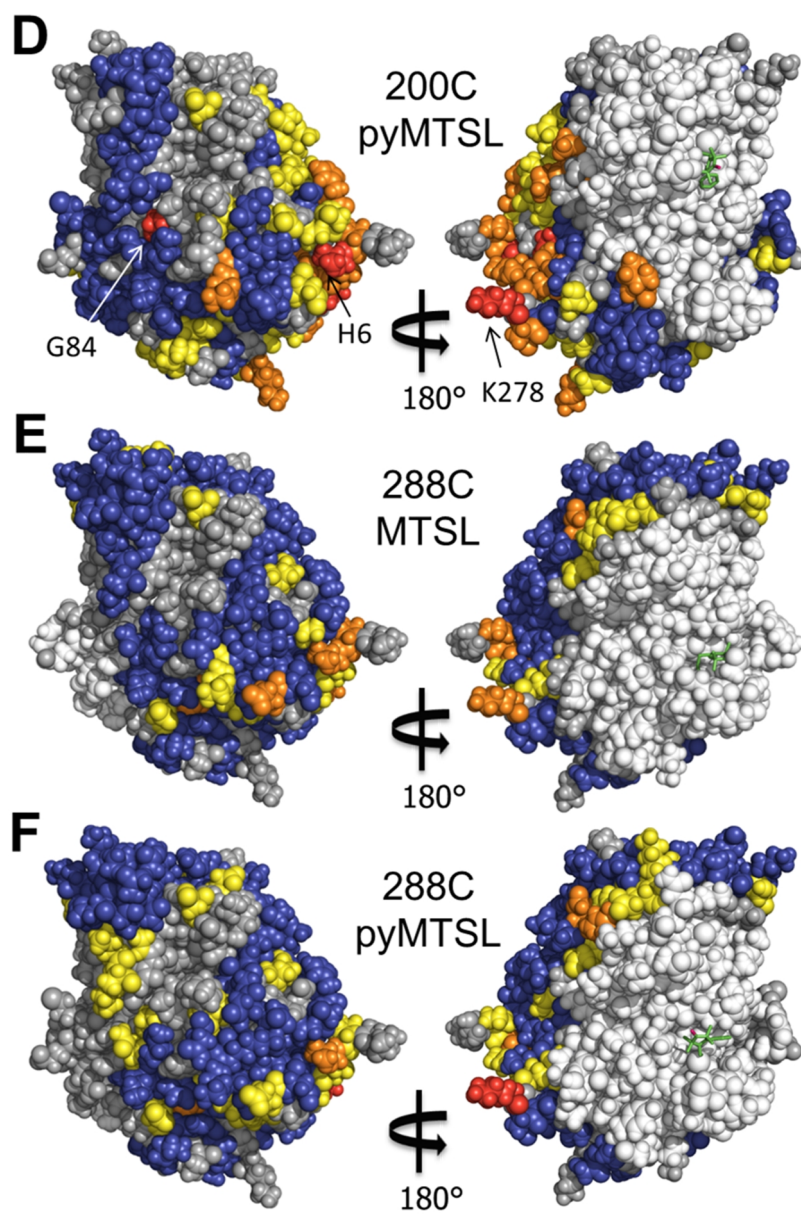


FIGURE 6.5 Continued

The distance of the CcP residues from the protein's surface was calculated by determining the closest distance of that residue to an external bulk solvent (water), excluding a water radius of 2.6 Å.^(322, 323) When comparing the intermolecular PRE data to the distance of the amides from the surface of CcP, PRE are not necessarily expected for all amides near the surface of CcP but they are not expected for deeply buried residues. This is indeed what is seen for most of the intermolecular PRE data which do not show effects for residues deeper than 8 Å. This is demonstrated particularly nicely for residues 50-80 with the SL attached at position 38C or 200C (Figure 6.3 A-D). For 288C, two moderately strong PRE are observed for residues 107 and 108 (17.7 Å and 17.1 Å for MTSL and 17.1 Å and 19.1 Å for pyMTSL, respectively) and those residues are 10.6 Å and 10.8 Å from the surface, respectively (Figure 6.3 E-F). Although large, these distances are not inconsistent with an intermolecular PRE. Conversely, less of a correlation between residue depth and PRE strength is observed for the intramolecular PRE. For example, very strong PRE are observed for all the residues between 195-207 and 248-262 for 200C although the depth of those residues from the surface ranges between 3.7-10.2 Å. The same holds for residues 255-263 for 288C. Several intramolecular PRE are also observed for residues > 8 Å from the surface such as residue 230 for 38C and residues 115 and 170 for 200C.

Interestingly, when Cc was added to the sample, the intermolecular PRE decreased. The dissociation constant, K_D , for the interaction between Cc and CcP is 5 μM,^(60, 140, 183) while the value for the CcP-CcP interaction would be expected to fall in the mM range.⁽³⁰³⁻³⁰⁵⁾ Therefore, assuming the CcP-CcP interaction occurs via the stereo-specific binding interface for Cc, the addition of Cc to the sample should reduce the observed intermolecular PRE for CcP. This can be seen by comparing the $R_{2\text{para}}$ values for samples containing only CcP or a 1:1 mixture with Cc (Figure 6.4). Note that $R_{2\text{para}}$ values greater than 128 s⁻¹ cannot be measured accurately as the peaks were entirely broadened out in the NMR spectra and are therefore estimates (peak intensity was set at twice the standard deviation of the noise level). Upon addition of Cc, the observed $R_{2\text{para}}$ values decreased dramatically for most of the residues that were further than 24 Å away from the nitroxide oxygen atom. This confirmed that the PRE for these residues were indeed caused by intermolecular interactions between different CcP molecules. Furthermore, that fact that these effects were diminished upon addition of Cc demonstrated that the CcP-CcP interaction was much weaker than that between Cc and CcP.

To visualize the localization of the interaction between the CcP molecules, the PRE effects were mapped on a surface model of CcP (Figure 6.5). Intramolecular PRE are expected within 24 Å of the nitroxide oxygen atom so for clarity these residues were excluded from the

PRE map (coloured white in figure 6.5). The intermolecular PRE show a clear localization to the “back” of CcP (relative to the stereo-specific binding site), centred around K278 (Figure 6.5 C & F). The only exception is K123 which is located closer to the “front” (Figure 6.5 C & D). A full list of the residues strongly affected by intermolecular PRE (coloured red in figure 6.5) is given in Table 6.1.

TABLE 6.1 CcP residues strongly affected ($R_{2\text{ para}} \geq 100 \text{ s}^{-1}$) by intermolecular PRE caused by spin labels attached at positions 38C, 200C or 288C. These residues are coloured red in Figure 6.5.

Position	MTSL	pyMTSL
38C	T4, L5, G273, K278, K268	T4, V7, G84, K278
200C	T4, F5, H6, V7, V10, K123, K268, G273, F276, K278	T4, F5, H6, V7, V10, K123, K268, G273, F276, K278
288C	T4, V7, G84, G273, K278	K278

Four of these CcP residues (H6, K123, K268, K278) are positively charged at pH 6 and residue T4 is located beside K3, which also showed a moderately strong (PRE Figure 6.2). The remaining residues are the non-polar amino acids leucine, valine, phenylalanine or glycine. Noticeably none of them are negatively charged. Negatively charged residues are unlikely to interact with CcP, which itself is highly negatively charged ($pI = 4.5$)⁽¹³⁵⁾ (Figure 6.6).

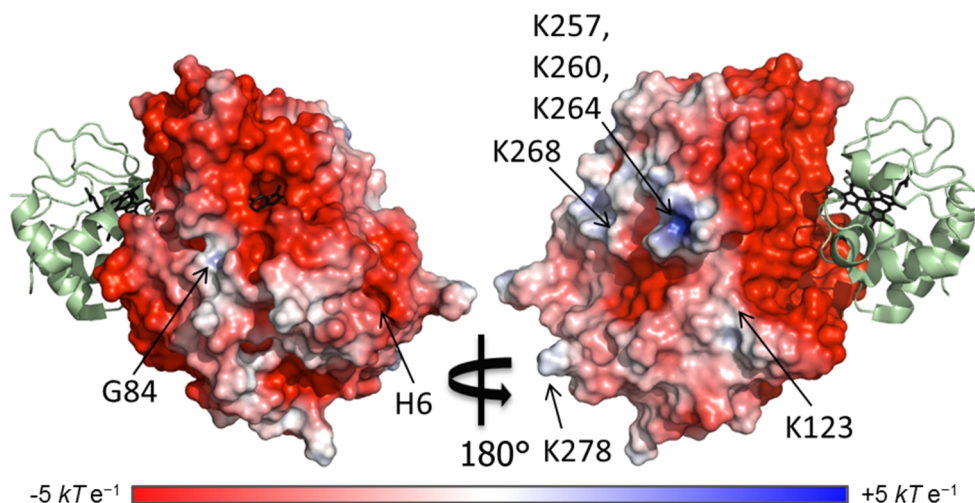


FIGURE 6.6 Map of the electrostatic potential generated for CcP (Cc is shown as green ribbon to identify the stereo-specific binding interface); the haem group shown in black sticks (PDB-entry 2PCC).⁽¹⁶⁹⁾ The potential isocontours range from -5 kT e^{-1} (red) to $+5 \text{ kT e}^{-1}$ (blue) and were calculated using ABPS.⁽³²⁴⁾

The electrostatic potential map for CcP shows the well-known large negative patch around the stereo-specific binding interface,⁽³⁸⁾ which interacts with the highly positively charged Cc.⁽¹²⁰⁾ There are also smaller negative patches on the sides and back of CcP. Noticeably, very few positive areas are present. The few positive areas that are present include the lysine residues that experience large PRE (84, 123, 268 & 278). This shows that although the CcP-CcP interaction appears to mainly be centered around K278, there are multiple sites of interaction. Interestingly, there is also a positive patch, comprised of lysines 257, 260 and 264, where an interaction would be expected but those residues do not experience any PRE (Figure 6.5). Unfortunately, intermolecular PRE data for these residues was only available for position 38C as these residues are within the intramolecular PRE range for positions 200C and 288C. Nevertheless, this shows that the interaction between CcP molecules is not merely localized around every positive patch. Since these residues are located next to the very negative stereo-specific binding, electrostatic repulsion could be the cause. It could also be that specificity is driven by more than just electrostatics. Given the number of non-polar amino acids that also experienced strong PRE (Table 6.1), perhaps hydrophobic interactions are also playing a role, although no obvious hydrophobic interaction patch was identified.

PRE distances can be used as restraints during protein-protein docking to build a structural model of a protein complex.⁽⁷⁷⁾ However, in the present study, the complex could not be modelled because of the symmetric nature of the SL interactions, which made ensemble docking complicated to perform. In order to model the CcP-CcP complex, the intermolecular PRE measurements could be repeated using non-isotopically labelled CcP single mutants that are spin labelled in a mixture with WT ¹⁵N isotope labelled CcP. In this way, only the intermolecular PRE will be observed and thus the derived distances can be used for protein docking. Nevertheless, the present study demonstrates the extreme sensitivity of PRE for minor states as the level of detail the PRE data provided was much greater than that from the line broadening in the NMR spectra. Furthermore, this evidence for an ultra-weak CcP-CcP complex may have implications for other studies, particularly when investigating weak effects, and should be taken into account when designing future experiments.

CHAPTER 7

Combining multiple types of paramagnetic NMR data to study the cytochrome *c* – cytochrome *c* peroxidase encounter complex

ABSTRACT

The interaction between Cc and CcP is transient and highly dynamic making this complex an ideal candidate for studying the encounter complex. Previous pNMR studies demonstrated that the orientation of the complex found in the crystal structure, the stereospecific state, accounted for only 70% of the lifetime of the complex in solution and the remainder was spent in a loosely associated encounter state. However, this model has previously only been fit to PRE data which, although exquisitely sensitive to minor states, is not ideal for determining the relative orientations of proteins within a complex. Here we use a CLaNP-5 probe coordinated to various Ln³⁺ ions to generate PRE as well as PCS and RDC, which do provide orientation information. Unfortunately, we find a poor fit between the theoretical and experimental data due to disruption of complex formation caused by steric and electrostatic repulsion between Cc and CLaNP-5. For two of the mutants, this likely results in a larger population in the encounter state as well as a larger spread in the number of Cc orientations than is accounted for in the existing model. Nevertheless, developing a model of the encounter complex using all three types of paramagnetic data remains an important goal for understanding how the complex behaves in solution and, therefore, this work should be repeated with CLaNP-7 or other less charged probes.

INTRODUCTION

The traditional view of protein complex formation as a simple two-step process has become outdated with the discovery in recent decades of the encounter complex. The encounter complex is a highly dynamic transition state along the binding trajectory in which the proteins pre-orient themselves for stereo-specific complex formation.⁽³⁾ The formation of the encounter complex is driven by long-range interactions such as electrostatic and desolvation forces.⁽⁴⁾ As the proteins sample each-others surfaces, the number of contacts between the two increases until optimal binding geometry is reached allowing short range interactions to form, such as vdW, hydrogen-bonding, hydrophobic contacts and specific salt bridges, and the complex proceeds to the tightly bound, stereo-specific state.⁽³⁾ ET complexes often make use of this intermediate, loosely associated encounter state to counterbalance their requirement for a high turnover rate while at the same time achieving a specific enough interaction to allow ET to occur.⁽⁵⁾

The Cc-CcP complex is a well-studied model complex for ET. In the cell, Cc transfers two reducing equivalents to CcP, which are then used to catalyse the reduction of hydrogen peroxide to water.⁽¹³⁵⁾ The crystal structure of this complex has been available since 1992 and shows the relative orientation of Cc on CcP in the stereo-specific state.⁽¹⁶⁹⁾ In 2006, Volkov *et al.* attached MTSL to the surface of CcP and observed the resulting PRE effects in the NMR spectra of Cc. The orientation of the complex in solution was found to be close to that of the crystal structure but the stereo-specific state only accounted for 70% of the lifetime of the complex, with the remaining 30% spent in an encounter state.⁽¹⁾ This was confirmed in a more extensive paramagnetic (pNMR) study by Bashir *et al.*, in 2010.⁽⁸⁾ Later it was also found that the population in the encounter state could be shifted to as low as 10% or as high as 80% with targeted point mutations in the binding interface.⁽⁹⁾

In 2014, Schilder *et al.*, observed the complex from the other side by attaching MTSL on the surface of Cc and observed the resulting PRE effects in the NMR spectra of CcP (Chapter 4).⁽¹⁴⁰⁾ The new data fit the previous 30% encounter/70% stereo-specific (30:70) model^(1, 8) well for most CcP residues but some discrepancies were found. This highlighted the importance of obtaining a comprehensive data set when working with pNMR data.⁽¹⁴⁰⁾ All pNMR effects (PRE, RDC, PCS) represent an average over all the conformations present in the sample when there is fast interconversion between them on their respective NMR timescales. This makes visualization of the complex an *ill-posed inverse problem*,^(273, 274) in which many ensembles of solutions can be found to match the observed data.^(3, 8, 10, 11, 13, 14, 25, 34-36)

The only way to exclude false solutions is by obtaining additional data sets. This can be done using tags at different locations, using different types of paramagnetic centres or using different types of paramagnetic effects.^(32, 33, 275-277)

In 2014, Volkov *et al.* extended the available PRE data set by using EDTA-Mn²⁺ tags attached on the surface of Cc and observing the PRE effects in the NMR spectra of CcP while altering the oxidation state of the two proteins.⁽²⁰⁴⁾ When using RS CcP in the presence of ferric Cc (Cc^{ox}), the same as was done in previous studies, they found that data fit well to the 30:70 model.^(1, 8) Furthermore, they also measured PCS in the CcP spectra generated by the LS Fe³⁺ atom in ferrous Cc (Cc^{red}). Interestingly, for CcP(CN⁻)-Cc^{red}, they found a slightly better match between the experimental and back-predicted PRE and PCS data when the population of the encounter state was increased to 40%. This was likely due to subtle differences in the binding of Cc^{red} or Cc^{ox} as overall little difference was found between CcP(RS)-Cc^{ox} and CcP(CN⁻)-Cc^{red} complexes.⁽²⁰⁴⁾ Nevertheless, this demonstrates that additional information can be obtained by combining PCS and PRE data.

Although PRE are extremely sensitive to minor states,⁽⁵⁹⁾ they are not ideal for determining relative orientations within a protein complex (Chapter 5).⁽³²¹⁾ Much better results can be obtained if PRE (distance information) are combined with PCS (distance and orientation information) and RDC (orientation information).^(14, 31, 80, 104-108, 299-301) Furthermore, the quality of pNMR results are also highly dependent on tag mobility as unwanted movement of the paramagnetic centre can lead to excess signal averaging and therefore an underestimation of the true signal.⁽³²⁵⁾ This can be limited by using the highly immobile, double-armed CLaNP-5-Ln³⁺.^(85, 102, 326) These probes will chelate any Ln³⁺ ion allowing for the fine-tuning of desired paramagnetic effects. For PCS and RDC, lanthanoids with anisotropic $\Delta\chi$ -tensors such as holmium (Ho³⁺) or thulium (Tm³⁺) can be used, which produce measurable effects within approximately 10-40 Å⁽⁷²⁾ and 15-70 Å,^(301, 327) respectively. For PRE, an isotropic $\Delta\chi$ -tensor is required as is the case for gadolinium (Gd³⁺) which can produce measurable effects within approximately 20-45 Å.⁽³⁰¹⁾ In combination, these ranges will allow for complete coverage of the entire Cc-CcP complex. Finally, as a diamagnetic control, lutetium (Lu³⁺) is available.

In this study, we attach CLaNP-5-Ln³⁺ probes at four positions on the surface of CcP. First, we measured the CSP and determine the binding coefficients for the tagged protein. Then, using intramolecular PCS data, we determine the $\Delta\chi$ -tensors and the positions of the paramagnetic centres. With this information, we determine the intermolecular PCS, RDC and PRE effects which are then compared to the existing 30:70 model of the complex that has

previously been fit to PRE data.^(1, 8) Unfortunately, we find that CLaNP-5 disrupts complex formation to some degree for all mutants and also appears to increase the population of the encounter complex for two of the mutants. This results in a poor fit between the experimental data and the theoretical models since the 30:70 model likely underestimates the spread of the Cc orientations sampled by this increased encounter state population. The strong positive charge of the probe is likely a contributing factor to this effect and therefore CLaNP-5 is not ideal for studying proteins whose interaction is driven predominately by electrostatics.^(169, 328) Future studies should be done with CLaNP-7 or other less charged probes.⁽³²⁹⁾

MATERIALS & METHODS

PROTEIN PURIFICATION

The pET28aCcP plasmid⁽¹⁾ was modified to contain double cysteine mutations at positions 24/28, 87/90, 212/214 and 249/253 using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) by Anneloes Blok. The CcP mutants were expressed and purified as published previously (Chapter 3).^(139, 140, 142) A pUC19 based plasmid containing the *S. cerevisiae* iso-1-cytochrome *c* gene was used to express and purify Cc as described previously.^(124, 125)

Ln³⁺-CLaNP-5 ATTACHMENT

The CLaNP-5-Ln³⁺ tag containing either Lu³⁺, Ho³⁺, Tm³⁺ or Gd³⁺ as the Ln³⁺ was prepared as described previously.⁽¹⁰²⁾ Approximately 20 mg of a double-cysteine CcP mutant was incubated for 1 hr at room temperature with 5 mM DTT in NMR buffer (20 mM NaPi, 100 mM NaCl) at pH 7. Meanwhile, 15 mL of NMR buffer was de-gassed and then bubbled with argon for 1 hr. The DTT was removed from the protein solution using a PD-10 column; as the protein eluted from the column, it was dripped into the degassed NMR buffer containing 5 molar equivalents of Ln³⁺-CLaNP-5. The sample was incubated at 4 °C for at least 1 hr. Unlabeled protein, protein oligomers and surplus of Ln³⁺-CLaNP-5 were removed using a gel filtration column (Superdex 75 10/300 GL, GE Healthcare). The tagging efficiency was determined to be ~90% by comparing the intensities of diamagnetic residual peaks with PCS shifted peaks in the NMR spectra of CLaNP-5-Ho³⁺ and CLaNP-5-Tm³⁺ CcP.

NMR SPECTROSCOPY

NMR TITRATIONS

The binding constants were obtained for each CLaNP-5-Lu³⁺ tagged CcP double cysteine mutant by titrating a 5.2 mM stock of WT Cc into 400 μ M ¹⁵N labelled CcP as described previously (Chapter 4).⁽¹⁴⁰⁾ The average CSP ($\Delta\delta_{\text{avg}}$) were also derived and extrapolated to 100% bound as described previously.⁽²⁸⁰⁾

PARAMAGNETIC NMR

The intramolecular paramagnetic effects for CLaNP-5-Lu³⁺, CLaNP-5-Ho³⁺ and CLaNP-5-Tm³⁺ tagged ¹⁵N-CcP double cysteine mutants were recorded using 2D BEST-TROSY-HSQC experiments ⁽²⁷⁹⁾ on a Bruker AVIII HD spectrometer equipped with a ¹H[¹³C/¹⁵N] TCI-cryoprobe operating at a Larmor frequency of 850 MHz at 293 K with 1024 and 100 complex points in the ¹H and ¹⁵N dimensions respectively. The NMR buffer contained 20 mM NaPi, 100 mM NaCl, 6% D₂O, pH 6.0. NMR spectra for mutant 87/90C were also recorded at 303 K for comparison. The intermolecular paramagnetic effects for CLaNP-5-Lu³⁺, CLaNP-5-Ho³⁺, CLaNP-5-Tm³⁺ and CLaNP-5-Gd³⁺ tagged CcP double cysteine mutants on ¹⁵N Cc WT were recorded using 2D HSQC⁽³³⁰⁾ experiments on a Bruker AVIII HD spectrometer equipped with a ¹H[¹³C/¹⁵N] TCI-cryoprobe operating at a Larmor frequency of 850 MHz at 303 K with 1024 and 128 complex points in the ¹H and ¹⁵N dimensions respectively. The intermolecular RDC was measured using IPAP-HSQC⁽³³¹⁾ experiments on the same spectrometer with 1024 and 128 complex points in the ¹H and ¹⁵N dimensions respectively. The NMR samples contained 300 μ M of each protein in 20 mM NaPi, 100 mM NaCl, 6% D₂O, pH 6.0. All data were processed using Topspin 3.2 (Bruker, Karlsruhe, Germany) and analysis was done using CCPN analysis 2.1.5.

DATA ANALYSIS

PCS AND RDC ANALYSIS

Using the intramolecular PCS data, the $\Delta\chi$ -tensors for CLaNP-5-Ho³⁺/Tm³⁺ tagged CcP were determined in Numbat⁽⁹⁶⁾ using the measured PCS values and the structure of CcP from the crystal structure of the complex (PDB entry 2PCC)⁽¹⁶⁹⁾ to which HN atoms were added using Xplor-NIH.^(282, 283) For the intermolecular PCS data, the measured PCS values were extrapolated to 100% bound Cc and the position of the Ln³⁺ ions were fixed to the

coordinates determined during the intramolecular $\Delta\chi$ -tensor calculations. The intermolecular $\Delta\chi$ -tensors were determined in Numbat⁽⁹⁶⁾ using the extrapolated PCS values and the structure of Cc from the crystal structure of the complex (PDB entry 2PCC).⁽¹⁶⁹⁾ The Q-values for the intramolecular PCS were calculated according to equation 5.2 as described previously (Chapter 5).^(8, 321) The RDCs were measured directly from the IPAP-HSQC spectra (in ppm) and were first converted to Hz and then extrapolated to 100% bound Cc. An error of ± 6 Hz was extrapolated from the error in peak peaking. The theoretical values were predicted by Numbat⁽⁹⁶⁾ based on the PCS based intermolecular the $\Delta\chi$ -tensors.

Since no data were recorded for CLaNP-5-Ho³⁺ at positions 24/28 and 249/253, the $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values were estimated based on the ratios between these values for Ho³⁺ and Tm³⁺ at positions 87/90 and 212/214. The Ho³⁺ and Tm³⁺ $\Delta\chi$ -tensors were then used to calculate the theoretical paramagnetic effects (PCS and RDC) and the distances between the amide protons of Cc and the Ln³⁺ atom for the stereo-specific complex. The same was done for each of the 1701 orientations of Cc observed in the MC ensemble used to describe the encounter complex before.⁽⁸⁾ Orientations in the ensemble in which the lanthanoid approached any of the Cc C α atom to less than 6 Å were excluded, because they were taken to represent a penetration of Cc by the CLaNP-5-Lu³⁺. For mutants 24/28, 87/90, 212/214 and 249/253 this resulted in 2, 0, 13 and 3 exclusions out of 1701 orientations, respectively. To model the 30% encounter/70% stereo-specific complex, the back-calculated values for each were combined; r^{-6} averaging was used for combining the distances. Molecular manipulations and distance measurements were done using XPLOR-NIH.^(282, 283)

PRE ANALYSIS

The intensity ratio of the amide resonances in the spectra of the paramagnetic (Gd³⁺) and diamagnetic (Lu³⁺) samples (I_{para}/I_{dia}) was measured and normalized as described previously.⁽⁸⁾ The paramagnetic contribution to the transverse relaxation rate, $R_{2,para}$, was calculated as described previously (Chapter 4).^(8, 77, 140) The average $R_{2,dia}$ value was used with a large error margin for those amides for which an I_{para}/I_{dia} could be measured but for which the line width of the diamagnetic peak could not be obtained. For the amide peaks that disappeared in the paramagnetic spectrum, an upper limit for I_{para} was set to two standard deviations of the noise level of the spectrum.⁽¹⁴⁰⁾ The calculated $R_{2,para}$ values were then converted into distances as described previously,⁽⁸⁾ see equation 4.1, with a Cc fraction bound of 0.784, 0.727, 0.735 or 0.765 for positions 24/28, 87/90, 21/214 and 249/253, respectively. The value of τ_c for the complex was previously estimated to be 16 ns.^(1, 8)

RESULTS & DISCUSSION

EFFECTS OF PROBE ATTACHMENT

Four double cysteine mutations were generated on the surface of CcP in a large ring around the stereo-specific binding site: N24C/L28C, N87C/K90C, K212C/E214C and K249C/N253C. By attaching CLaNP-5-Ln³⁺ probes at these positions, the entire complex could be observed using pNMR (Figure 7.1). Previous pNMR studies observed strong PRE effects on CcP when SLs were attached at positions 38C, 200C and 288C.^(1, 8) Two of these double cysteine mutations are very close to the attachment sites in previous studies; 24/28 is close to 288C and 249/253 is close to 200C, while 38C falls roughly between 24/28 and 87/90 (Figure 7.1A).

The CLaNP-5-Ln³⁺ probe is relatively large and can therefore potentially cause steric interference during complex formation. Furthermore, the net charge of attached probe is +3 which is not ideal for a complex that is itself highly charged (CcP is negative and Cc is positive) and relies mainly on electrostatics to drive the interaction.^(169, 328) Therefore, WT Cc was titrated into a solution of CLaNP-5-Lu³⁺ tagged ¹⁵N-CcP and the CSP caused by binding were measured. Several resonances shifted in the spectra indicating a fast exchange binding process, as is expected for the Cc-CcP complex. The binding constants (K_B) were determined by fitting the CSP curves for several amide resonances to a 1:1 binding model as described previously (Chapter 4)⁽¹⁴⁰⁾ (Figure 7.2). For the WT complex, the K_B is $2 \times 10^5 \text{ M}^{-1}$.^(60, 140, 183) The calculated K_B values for CLaNP-5-Ln³⁺-CcP range from $1.0 \times 10^5 \text{ M}^{-1}$ for position 24/28 to $0.5 \times 10^5 \text{ M}^{-1}$ for position 87/90 indicating that the presence of the probe is interfering somewhat with complex formation either through steric interferences or electrostatic repulsion. This difference in binding is taken into account during data analysis since all pNMR effects are extrapolated to 100% bound. The range in the magnitude of $\Delta\delta_{\text{avg}}$ is indicative of the degree of dynamics in the complex and therefore the relative amount of the complex in the encounter state.^(9, 15, 27, 332) The $\Delta\delta_{\text{avg}}$ ranges from approximately -0.75 to 0.62 for position 24/28. For positions 212/214 and 249/253 the $\Delta\delta_{\text{avg}}$ ranges from approximately -0.75 to 0.50 indicating perhaps a slight increase in the encounter state population but the strongest effect is seen for position 87/90 with a range from only -0.29 to 0.39. Therefore, it is possible that the populations of stereo-specific and encounter state may be shifted and this will need to be taken into consideration when comparing the experimental data to the 30:70 model.^(1, 8)

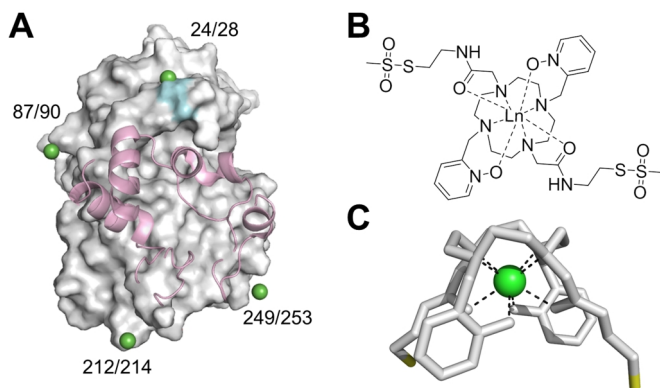


FIGURE 7.1 **A**) Experimentally determined locations of Ln^{3+} ion (green sphere) within the CLaNP-5 probe attached to double cysteine mutations at positions 24/28, 87/90, 212/214 and 249/253 on the surface of CcP (grey; C-terminus in light blue; N-terminus on the back-side) with respect to the binding site of Cc (semi-transparent pink cartoon) (PDB-entry 2PCC).⁽¹⁶⁹⁾ **B**) Molecular structure of CLaNP-5 with a coordinated lanthanoid ion.⁽¹⁰²⁾ **C**) 3D model of CLaNP-5 attached via two sulphur atoms (yellow) and containing a coordinated lanthanoid ion (green sphere).⁽¹⁰²⁾

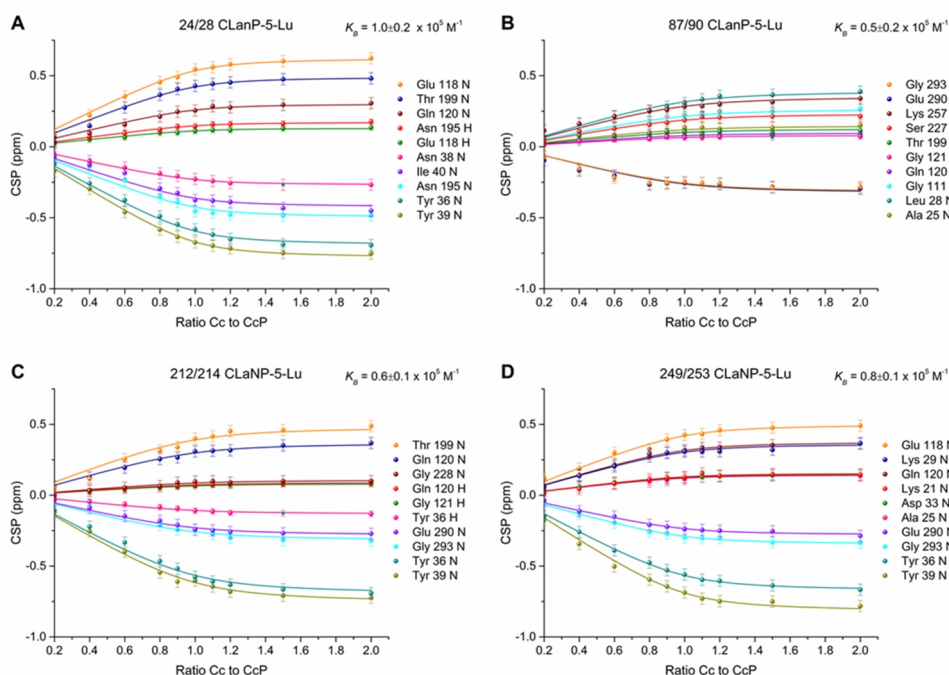


FIGURE 7.2 Chemical shift perturbations for selected CcP residues in the ^1H or ^{15}N dimension during titration with WT Cc. The titrations were done for CLaNP-5- Lu^{3+} CcP attached at positions 24/28 (**A**), 87/90 (**B**), 212/214 (**C**) or 249/253 (**D**). The curves were fitted globally to a 1:1 binding model and the solid lines show the best fit when using a shared K_B value. These experiments were done in 20 mM NaPi , 100 mM NaCl (pH 6.0) at 293 K.

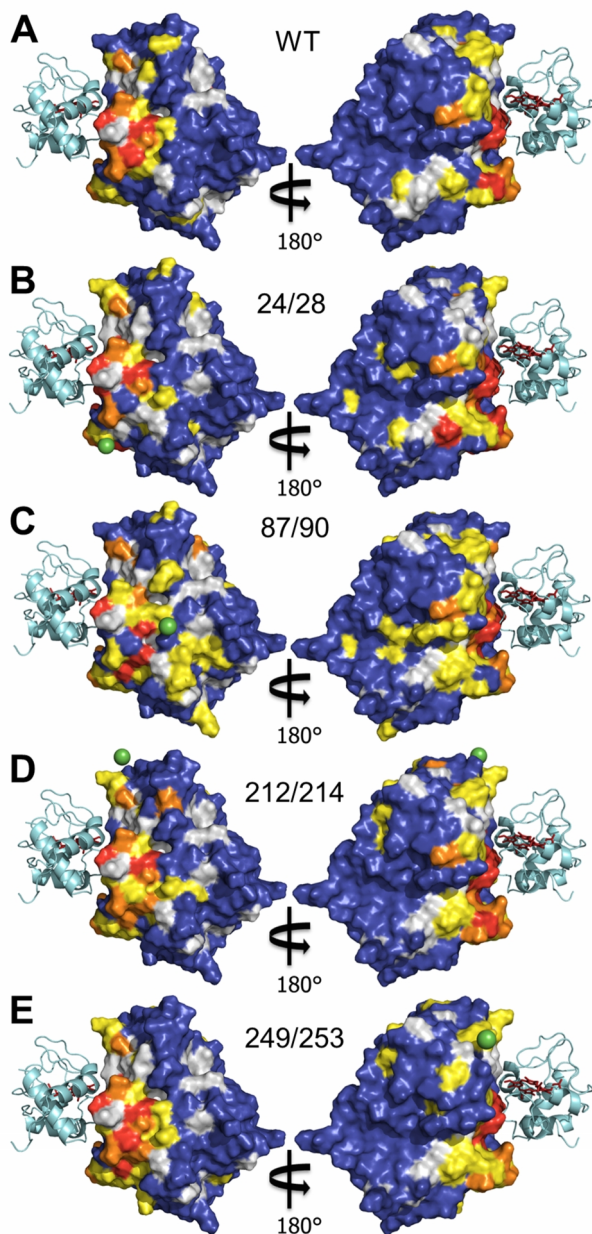


FIGURE 7.3 Chemical shift perturbation map for WT Cc with ^{15}N -CcP C128A⁽¹⁴⁰⁾ (**A**) or with CLaNP-5- Lu^{3+} attached at positions 24/28 (**B**), 87/90 (**C**), 212/214 (**D**) or 249/253 (**E**) with the lanthanoid ion shown as a green sphere. The CSP are colour coded on a surface model of CcP in the stereo-specific complex (PDB-entry 2PCC).⁽¹⁶⁹⁾ Cc is shown in cyan ribbon with the haem group in red sticks. CSP were extrapolated to 100% bound CcP. Residues with $\Delta\delta_{\text{avg}} \geq 0.06$ ppm are red, 0.04-0.06 ppm are orange, 0.02-0.04 ppm are yellow, 0-0.02 ppm are blue and with no data are grey.

Using the calculated K_B , the average amide shifts, $\Delta\delta_{\text{avg}}$, were extrapolated to 100% bound CcP and mapped onto the surface of CcP (Figure 7.3). Despite the differences in the binding constants for CLaNP-5-Lu³⁺-CcP in comparison to WT, relatively minor differences are observed in the CSP maps. The strongest CSP are located at the stereo-specific binding interface for all double mutants as expected and only a few extra weak effects are observed toward the back of CcP. This effect is strongest for 87/90, which has the lowest K_B value of $0.5 \times 10^5 \text{ M}^{-1}$. But this effect is not directly correlated to the K_B values for the other double mutants, as 24/28 also shows more weak CSP on the back of CcP ($K_B=1.0 \times 10^5 \text{ M}^{-1}$) that do either 212/214 ($K_B=0.6 \times 10^5 \text{ M}^{-1}$) or 249/253 ($K_B=0.8 \times 10^5 \text{ M}^{-1}$). Nevertheless, these data indicate that the presence of the CLaNP-5 probe is interfering to some degree with formation of the complex.

INTRAMOLECULAR DATA ANALYSIS

Intramolecular PCS were measured for CLaNP-5-Ho³⁺ at positions 87/90 and 212/214 and for CLaNP-5-Tm³⁺ at all four positions and the $\Delta\chi$ -tensors were calculated (Table 7.1).

TABLE 7.1 Intramolecular PCS derived $\Delta\chi$ -tensor values for Ho³⁺ and Tm³⁺ tensors for CLaNP-5 bound double cysteine mutants of yeast CcP C128A calculated using Nubat(96) with PCS obtained from a [¹⁵N,¹H]-BEST-TROSY-HSQC spectrum taken at 20 °C and the CcP crystal structure (PDB entry 2PCC(169)). The values in italics are estimates (see description in materials and methods).

Position	Ho ³⁺ $\Delta\chi_{\text{ax}}$ (10 ⁻³² m ³)	Ho ³⁺ $\Delta\chi_{\text{rh}}$ (10 ⁻³² m ³)	Tm ³⁺ $\Delta\chi_{\text{ax}}$ (10 ⁻³² m ³)	Tm ³⁺ $\Delta\chi_{\text{rh}}$ (10 ⁻³² m ³)
24/28	-24.7 ± 1.3	-4.4 ± 0.5	49.4 ± 0.3	16.7 ± 0.8
87/90	-28.0 ± 0.5	-9.5 ± 0.2	57.6 ± 1.1	36.0 ± 0.7
87/90 (30 °C)	-29.0 ± 0.9	-8.8 ± 0.4	57.6 ± 1.3	36.5 ± 0.9
212/214	29.6 ± 1.9	9.4 ± 0.8	54.7 ± 2.8	40.2 ± 2.1
249/253	-23.8 ± 1.2	-5.0 ± 0.5	47.6 ± 1.1	18.8 ± 1.3

The resulting $\Delta\chi$ -tensors had very similar $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ values at the different positions. For Ho³⁺, $\Delta\chi_{\text{ax}}$ was approximately $-28.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}}$ was approximately $9.4 \times 10^{-32} \text{ m}^3$ which was slightly larger (although in similar proportion) to the theoretical values of $18.5 \times 10^{-32} \text{ m}^3$ and $5.8 \times 10^{-32} \text{ m}^3$, respectively.(61) Note that for positions 24/28 and 249/253, no data for Ho³⁺ was recorded so the values listed are estimates based on the experimentally determined values for the other two positions (see materials and methods). It is important to note that although using these estimates to fit the intermolecular PCS and PRE results in a reasonable fit to the experimental data (discussed below), these estimates do not take into account any

differences in the orientation of the Ho^{3+} $\Delta\chi$ -tensor in comparison to that of Tm^{3+} and so caution must be used when interpreting any results based on these estimates.

The $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ values for Tm^{3+} were much larger than those for Ho^{3+} , as expected, but they also showed more variation. For positions 24/28 and 249/253, the values for $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ were very similar at approximately $48 \times 10^{-32} \text{ m}^3$ and $18 \times 10^{-32} \text{ m}^3$, respectively, while the values for positions 87/90 and 212/214 were approximately $57.5 \times 10^{-32} \text{ m}^3$ and $35 \times 10^{-32} \text{ m}^3$, respectively. The $\Delta\chi_{\text{ax}}$ values were similar to previously reported for CLaNP-5- Tm^{3+} ($\Delta\chi_{\text{ax}}=55.3 \times 10^{-32} \text{ m}^3$; $\Delta\chi_{\text{rh}}=6.9 \times 10^{-32} \text{ m}^3$)⁽¹⁰²⁾ but all tensors showed significantly more rhombic character. However, values in which the proportion of $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ were roughly equal to each other have also been reported ($\Delta\chi_{\text{ax}}=21.9 \times 10^{-32} \text{ m}^3$; $\Delta\chi_{\text{rh}}=20.1 \times 10^{-32} \text{ m}^3$).⁽⁶¹⁾ These differences are likely due to small conformational changes in the CLaNP-5 structure, affecting the coordination to the Ln^{3+} , as a result of slightly different protein structures at each of the binding sites.

The backbone amide resonance assignments for CcP were obtained at 20 °C⁽¹⁴⁰⁾ but those available for Cc were obtained at 30 °C.⁽¹⁸³⁾ Generally it is not a problem to switch between these two temperatures when doing NMR experiments, depending on which of the binding partners are being studied. However, this study required that the intra- and intermolecular PCS/RDC data be comparable, which could potentially be problematic since the magnitude of an anisotropic $\Delta\chi$ -tensor is temperature dependent.⁽³³³⁾ For example, the PCS would be expected to be approximately 3% larger at 20 °C than at 30 °C.⁽³³⁴⁾ Of course, it would be possible to simply record all data at the same temperature but changing the temperature can shift the resonances in the spectra, making the assignments unclear. In order to assess how much difference a 10°C change in temperature would cause in the $\Delta\chi$ -tensor, the PCS were measured for CLaNP-5- $\text{Ho}^{3+}/\text{Tm}^{3+}$ at position 87/90 at both 20 °C and 30 °C (Table 7.1). At 30 °C, $\Delta\chi_{\text{rh}}$ was 7.2% smaller but the $\Delta\chi_{\text{ax}}$ was 3.7% larger for Ho^{3+} and the difference was only 1.0% for both values for Tm^{3+} . In fact, the values for both metals were the same within error, indicating that any affect from temperature is not likely significant in comparison to the error in calculating the $\Delta\chi$ -tensors.

For all the $\Delta\chi$ -tensors, the back-predicted PCS were compared to the experimentally determined PCS (Figure 7.4).

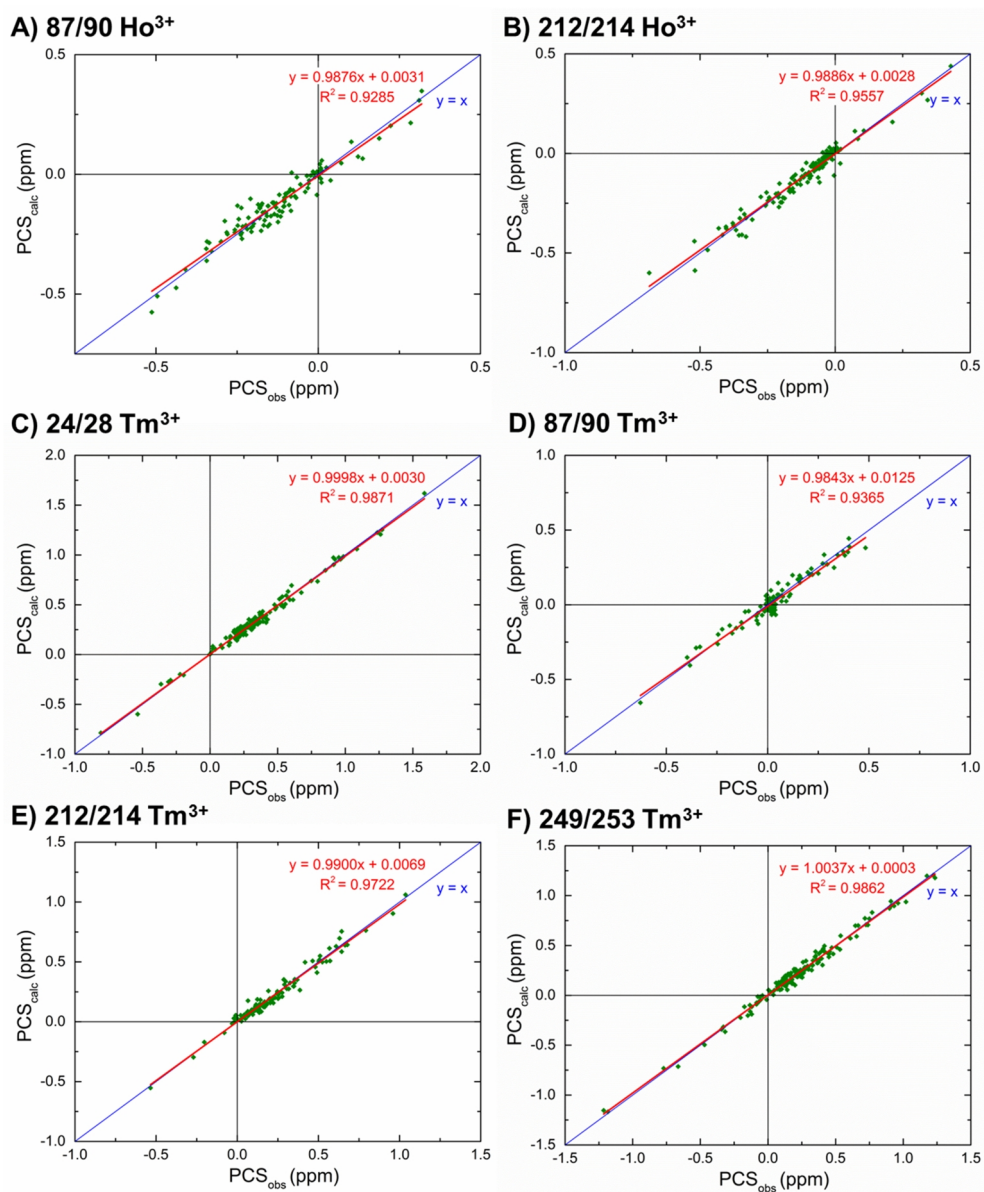


FIGURE 7.4 Back-calculated PCS plotted against experimentally observed PCS for CLaNP-5-Ho³⁺/Tm³⁺ attached to double cysteine mutants of CcP at positions 24/28 (**C**), 87/90 (**A**, **D**), 212/214 (**B**, **E**) or 249/253 (**F**). The linear best fit is shown in red and a $y=x$ line is shown in blue. The $\Delta\chi$ -tensor was calculated with Numbat⁽⁹⁾ using intramolecular PCS and the CcP structure in the crystal structure of the complex with Cc (PDB entry 2PCC⁽¹⁶⁹⁾).

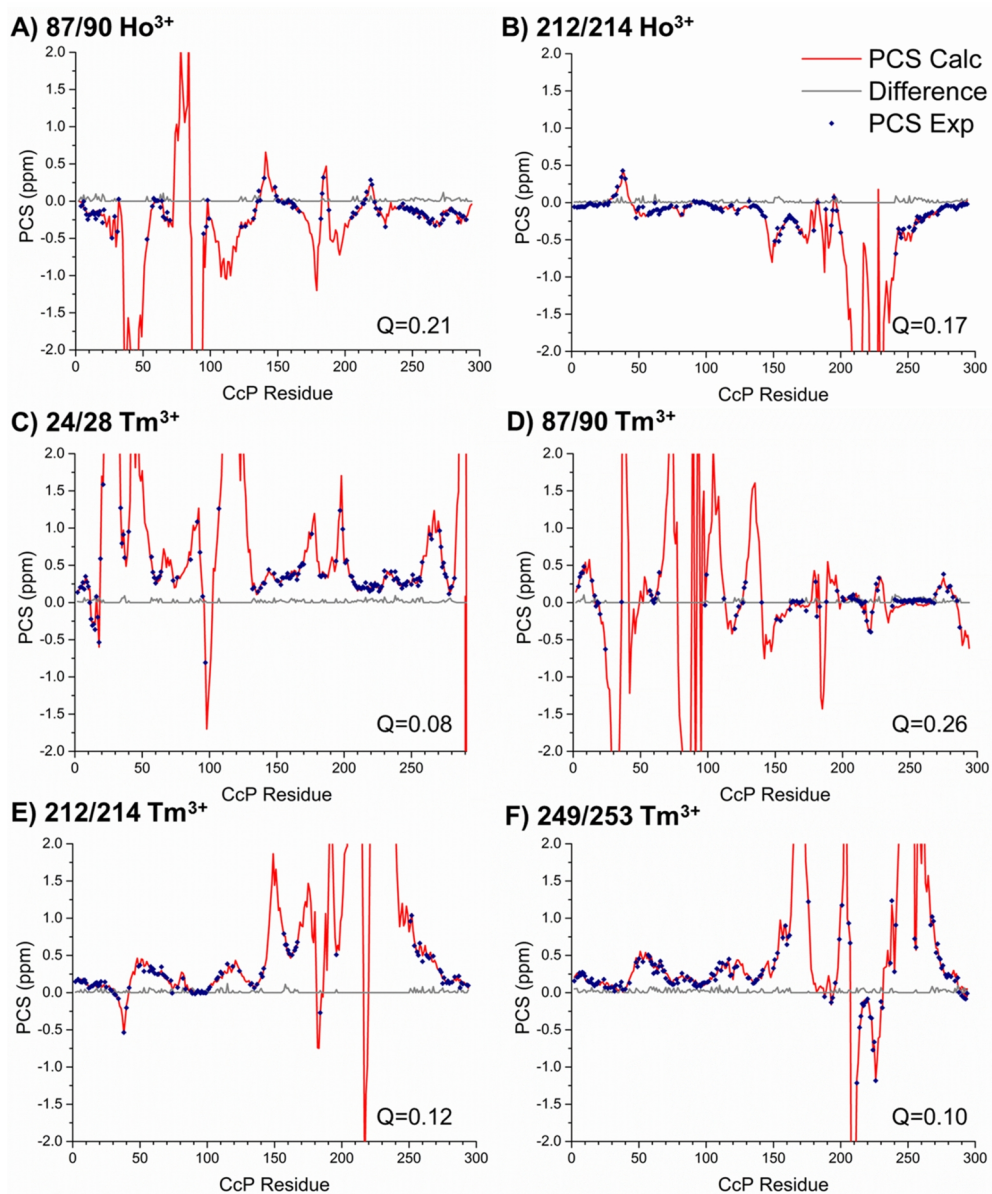


FIGURE 7.5 Comparison of the back-calculated (red line) and experimentally observed (blue dots) PCS on CcP amide hydrogen atoms for the Ho³⁺ and Tm³⁺ $\Delta\chi$ -tensors of CcP with CLaNP-5 attached to double cysteine mutants of CcP at positions 24/28 (**C**), 87/90 (**A**, **D**), 212/214 (**B**, **E**) or 249/253 (**F**). The difference between the two is plotted in grey. The average error for the PCS was ± 0.002 ppm. The $\Delta\chi$ -tensor was calculated with Numbat⁽⁹⁰⁾ using intramolecular PCS and the CcP orientation in the crystal structure of the complex with Cc (PDB entry 2PCC⁽¹⁶⁹⁾).

When the back-calculated and experimentally observed PCS are plotted against each other for all mutants, a good correlation is obtained.

Similar results were seen when the back-calculated and experimentally observed PCS were compared for all CcP residues (Figure 7.5). Overall, the fit between the back-calculated and experimentally observed intramolecular PCS was very good for all positions. There were small deviations in the PCS for some residues but they were spread across the sequence and no obvious areas of strong mismatch could be identified. As expected, the largest predicted PCS occurred closest to the respective attachment site for each CcP mutant. The position of the paramagnetic centre used during the subsequent intermolecular $\Delta\chi$ -tensor calculations was fixed at the position determined during these intramolecular $\Delta\chi$ -tensor calculations. These $\Delta\chi$ -tensors were also used to back-predict pNMR data for different models of the complex (discussed below).

INTERMOLECULAR PCS

Intermolecular PCS were measured for Cc in the presence of CLaNP-5-Ho³⁺/Tm³⁺ tagged CcP. As expected, there were fewer peaks in the paramagnetic spectra than the diamagnetic due to increased relaxation for nuclei close to the paramagnetic centre. Line-broadening due to chemical exchange may have also occurred. For a given nucleus in Cc, the CcP-bound and free forms will each have a characteristic resonance frequency. The magnitude of the difference between these frequencies, $\Delta\omega$, in comparison to the exchange rate, k_{ex} , defines the type of chemical exchange occurring; chemical exchange can be classified as fast ($k_{ex} \gg \Delta\omega$), intermediate ($k_{ex} \approx \Delta\omega$) or slow ($k_{ex} \ll \Delta\omega$).⁽³⁰⁹⁾ The Cc-CcP complex generally experiences fast-exchange on the NMR timescale and therefore each nucleus will appear as single peak in the spectrum, representing a weighted average of the two resonance frequencies. However, the use of strong paramagnetic centres, such as Tm³⁺, can potentially shift some nuclei into the intermediate or slow exchange regimes resulting in two separate peaks in the paramagnetic spectrum, as $\Delta\omega$ is the sum of the difference in the chemical shifts and the PCS. While the first is a relatively small effect, the PCS can be quite large, shifting the exchange regime for that nucleus to intermediate or even slow exchange.⁽³³⁵⁾ However, most observed PCS were ≤ 0.5 ppm, still in the fast or fast-intermediate exchange regime. The measured PCS values were extrapolated to 100% bound Cc and the $\Delta\chi$ -tensors were calculated (Table 7.2). The most notable difference when compared to the intramolecular $\Delta\chi$ -tensors (Table 7.1) is that all the values are much lower for the intermolecular $\Delta\chi$ -tensors (Table 7.2). The Cc-CcP

TABLE 7.2 Intermolecular PCS derived $\Delta\chi$ -tensor values for Ho^{3+} and Tm^{3+} tensors for CLaNP-5 bound double cysteine mutants of yeast CcP C128A calculated with Numbat⁽⁶⁰⁾ using PCS obtained from a $^{15}\text{N}, ^1\text{H}$ -BEST-TROSY-HSQC spectrum taken at 20 °C and the CcP crystal structure (PDB entry 2PCC⁽¹⁶⁹⁾).

Position	$\text{Ho}^{3+} \Delta\chi_{\text{ax}}$ (10^{-32} m^3)	$\text{Ho}^{3+} \Delta\chi_{\text{rh}}$ (10^{-32} m^3)	$\text{Tm}^{3+} \Delta\chi_{\text{ax}}$ (10^{-32} m^3)	$\text{Tm}^{3+} \Delta\chi_{\text{rh}}$ (10^{-32} m^3)
24/28	-10.7 ± 0.4	-3.3 ± 0.4	35.8 ± 1.4	8.7 ± 1.4
87/90	-11.1 ± 2.0	-1.9 ± 0.7	25.8 ± 1.8	8.8 ± 1.1
212/214	5.4 ± 2.4	2.3 ± 1.1	20.4 ± 1.5	3.1 ± 1.1
249/253	-14.5 ± 0.9	-5.3 ± 0.6	37.1 ± 0.9	2.5 ± 0.7

complex is known to partially exist in a highly dynamic encounter state that consists of multiple conformations in solution.^(1, 8) Since the observed PCS represent an average over all Cc conformations present in the sample, lower observed PCS and therefore lower $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ values are expected. The presence of the encounter complex contributes to a poorer fit between the predicted and experimentally observed PCS (Figure 7.6) because the predicted PCS values are based only on the stereo-specific orientation of the complex. However, the CSP data suggest that the presence of the tag is also disrupting complex formation to some degree and disturbances to the stereo-specific complex would also contribute to this poorer fit. The intermolecular PCS data also showed that it is beneficial to use both lanthanoids to obtain complete coverage of a protein complex; Ho^{3+} can provide information for nearby residues, whose peaks are completely broadened out by Tm^{3+} , while Tm^{3+} can provide long-distance information beyond the limit of Ho^{3+} .^(72, 301, 327)

In order to evaluate the fit to the 30:70 model^(1, 8) of the complex, the intermolecular PCS was plotted against back-predicted PCS values for a complex 100% in the stereo-specific state, 100% in the encounter state or a 70%/30% combination of the two using the $\Delta\chi$ -tensors obtained with intramolecular data (Table 7.1) (Figure 7.7). As mentioned above, the PCS represents an average of all conformations present in the sample; so nuclei sample a larger space in ensembles of conformation than in a stereo-specific structure, resulting in smaller PCS due to averaging of the $(3\cos^2\theta-1)$ factor (Equation 2.7). On the other hand, the PCS also depends on the distance between the observed nucleus and the paramagnetic centre (r^{-3}), so orientations that bring the nucleus closer to the paramagnetic centre may cause the average PCS to become larger. Both effects can be observed when comparing the back-predicted PCS for a 100% stereo-specific complex with those for a 100% encounter complex (Figure 7.7). Although none of the intermolecular PCS fit the back-calculated data for the 100% stereo-specific or the 30:70 model particularly well, the Tm^{3+} data for position 24/28 are the closest fit to the 30:70 model.

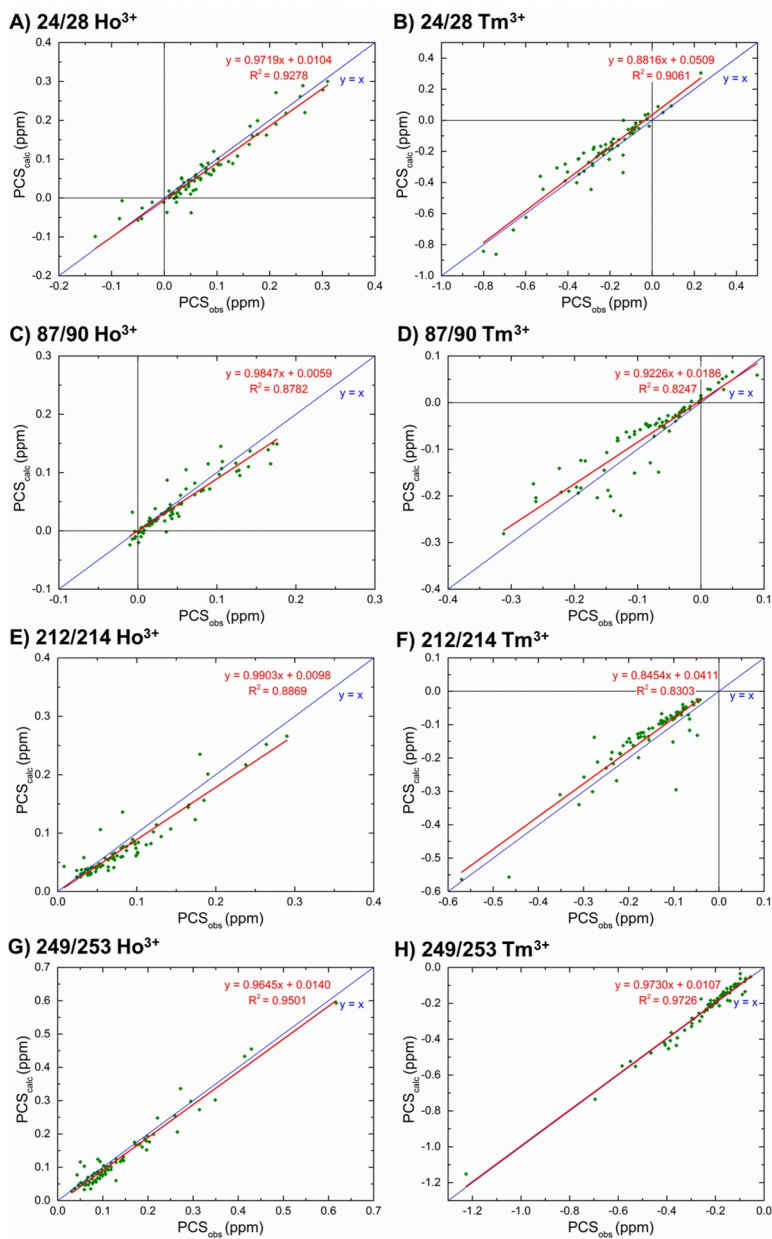


FIGURE 7.6 Back-calculated PCS plotted against experimentally observed PCS on Cc for CLaNP-5-Ho³⁺/Tm³⁺ attached to double cysteine mutants of CcP at positions 24/28 (**C**), 87/90 (**A, D**), 212/214 (**B, E**) or 249/253 (**F**). The linear best fit is shown in red and a $y=x$ line is shown in blue. The $\Delta\chi$ -tensor was calculated with Numbat⁽⁶⁶⁾ using intermolecular PCS the CcP orientation in the crystal structure of the complex with Cc (PDB entry 2PCC⁽¹⁶⁹⁾) and the Ln³⁺ position as determined from the intramolecular PCS.

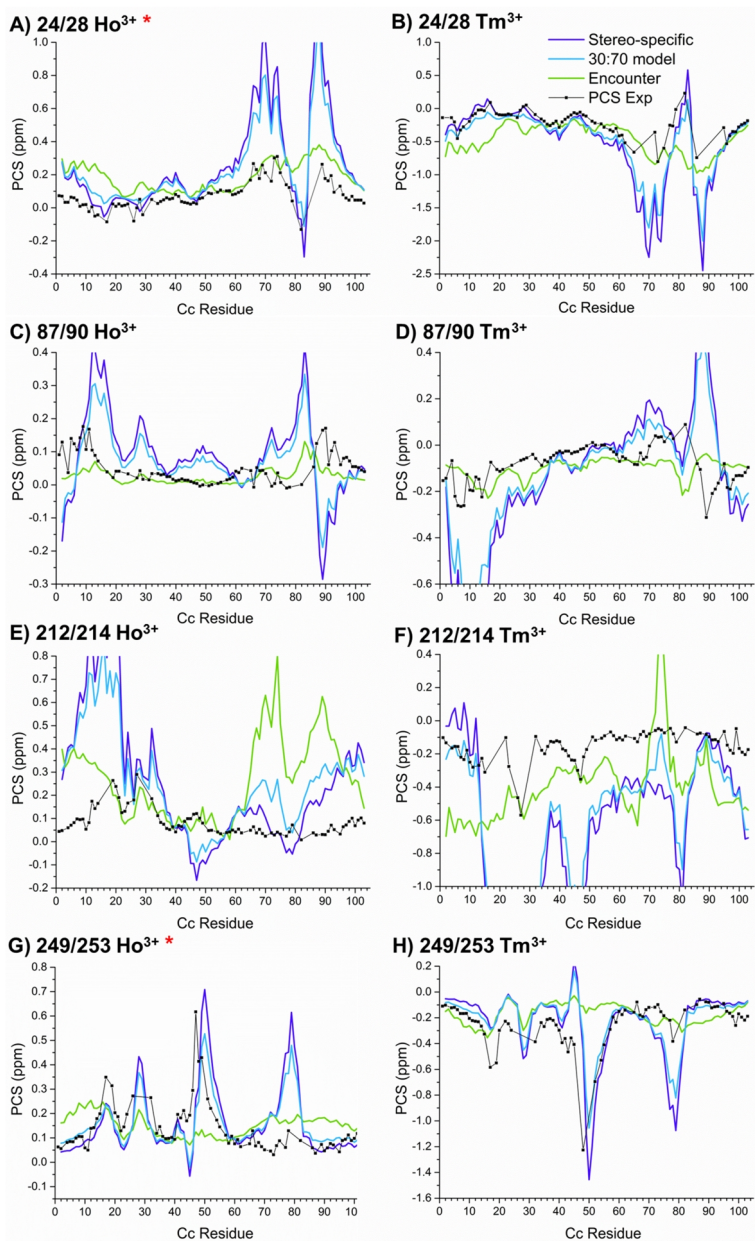


FIGURE 7.7 Analysis of intermolecular PCS data observed for Cc amide hydrogen atoms in a 1:1 complex with CLaNP-5- Ln^{3+} tagged CcP at positions 24/28 (**A, B**), 87/90 (**C, D**), 212/214 (**E, F**) or 249/253 (**G, H**). The experimental PCS (black dots) are compared to the average back-calculated PCS are also shown for 100% stereo-specific complex (purple line), 100% encounter complex (green line) or 30% encounter/70% stereo-specific complex (blue line). The Ho^{3+} $\Delta\chi$ -tensors used for back-calculation for positions 24/28 and 249/253 are estimates (red *). The average error for the PCS was ± 0.003 ppm.

The experimental PCS for the remaining positions show significant averaging in the values with effects that are mostly similar to or even weaker than those predicted for a complex 100% in the encounter state. The results suggest that the presence of the CLaNP-5 probe is disrupting stereo-specific complex formation and could also be causing an increase of the population in the encounter state. The very low experimental PCS values indicate that much more averaging is occurring in solution than is represented by any of these models, including the 100% encounter complex model, which could be explained by an under-sampling of the true breadth of Cc orientations present in the encounter complex.

INTERMOLECULAR RDC

RDC are relatively small and therefore difficult to measure accurately. Tm^{3+} can be used to generate large RDC, ≥ 20 Hz,⁽⁶¹⁾ but this comes at the cost of severe line-broadening for many peaks. However, RDC are not dependent on the distance between the probe and the observed nucleus so they can be very useful for obtaining orientation information over an entire protein complex. The intermolecular RDC were measured for Cc in the presence of CLaNP-5- Tm^{3+} CcP (Figure 7.8). Like PCS, RDC also represent an average of all conformations in solution and a large degree of averaging was observed for positions 87/90 and 212/214, with the experimental RDC values being of similar magnitude to those predicted for a 100% encounter complex. Interestingly, in contrast to positions 87/90 and 212/214 which fit the 100% encounter complex model best, the magnitude of the RDC for positions 24/28 and 249/253 was much closer to those predicted for the 30:70 model or the 100% stereo-specific complex. However, unlike the PCS, which can be very large, especially for Tm^{3+} , and easy to distinguish, the RDC can be small and difficult to measure with good precision. This is reflected in the large errors (± 6 Hz) for the measured RDC values, which makes it difficult to distinguish exactly which model the RDC fit best. For position 249/253, the PCS data also showed limited averaging and fit the 30:70 model much better than the 100% encounter complex model suggesting a difference in the encounter/stereo-specific populations for those positions compared to positions 87/90 and 212/214. The data for position 24/28 also fit the 30:70 well for most residues with limited averaging of the PCS; The fit is less good for residues 66-76 and 89-95 but the PCS for these nuclei were very large and only a few data points were could be measured.

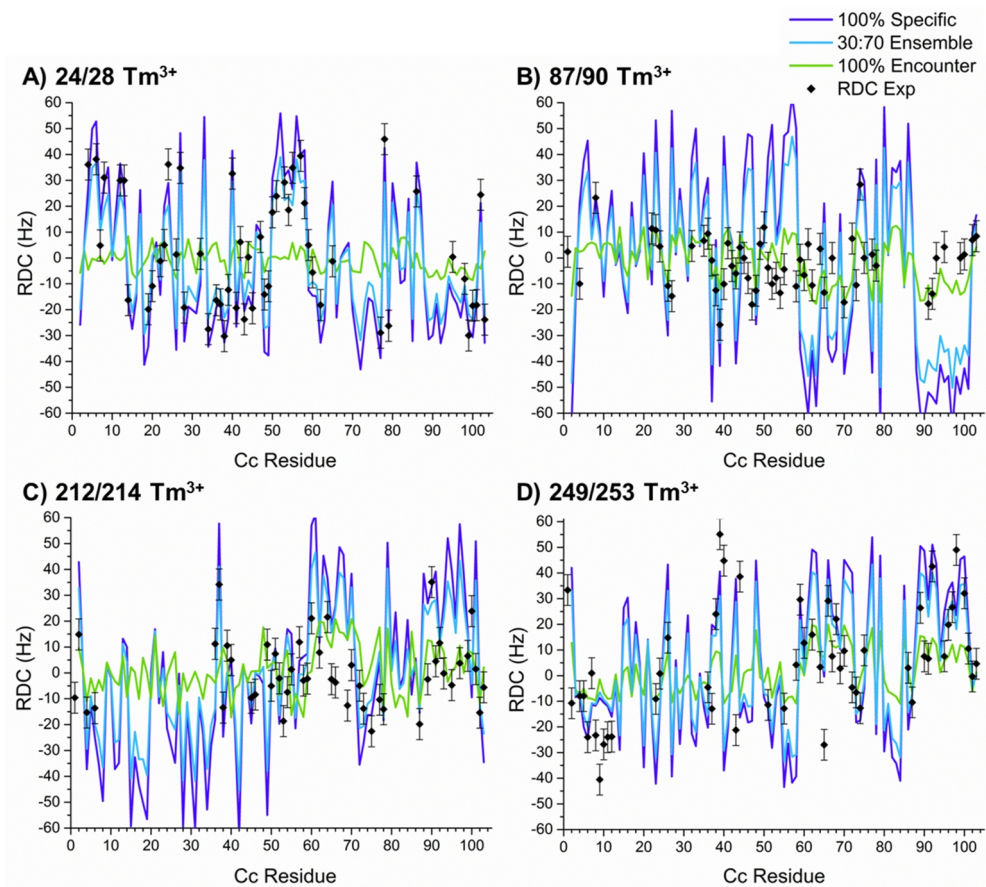


FIGURE 7.8 Analysis of intermolecular RDC data observed for Cc amide nitrogen atoms in a 1:1 complex with CLaNP-5-Ln³⁺ tagged CcP at positions 24/28 (**A, B**), 87/90 (**C, D**), 212/214 (**E, F**) or 249/253 (**G, H**). The experimental RDC (black dots with error of ± 6 Å) are compared to the average back-calculated RDC are also shown for 100% stereo-specific complex (purple line), 100% encounter complex (green line) or 30% encounter/70% stereo-specific complex (blue line).

PARAMAGNETIC RELAXATION ENHANCEMENT

Although RDC and PCS have the advantage of providing orientation information, PRE are exquisitely sensitive to minor states and therefore ideal for studying encounter complexes despite only providing distance information. Here we used CLaNP-5-Gd³⁺ attached at the four positions on CcP to generate long-range PRE effects on Cc. Three of the CLaNP-5 attachment sites used in this study are close to the sites used in previous studies: 24/28 is close to 288C, 249/253 is close to 200C and while 38C falls roughly between 24/28 and 87/90.^(1, 8)

The $I_{\text{para}}/I_{\text{dia}}$ ratios obtained with CLaNP-5-Gd³⁺ were similar to those observed previously at the nearby positions (Figure 7.9). Notably, position 24/28 experienced two areas of strong PRE near residues 69 and 89, which was also seen for 288C, around residues 75 and 82. For position 249/253, five areas experienced strong PRE (near residues 15, 29, 35, 50 and 80), four of which were similar to those seen for 200C (near residues 15, 25, 35 and 80).

The distance between the CLaNP-5 probes on CcP and the backside of Cc was less than 50 Å for all probe attachment sites. Due to the strength of the PRE generated by Gd³⁺, almost all Cc residues experienced some degree of relaxation enhancement resulting in essentially complete coverage of Cc, which can be seen in the PRE maps (Figure 7.10).

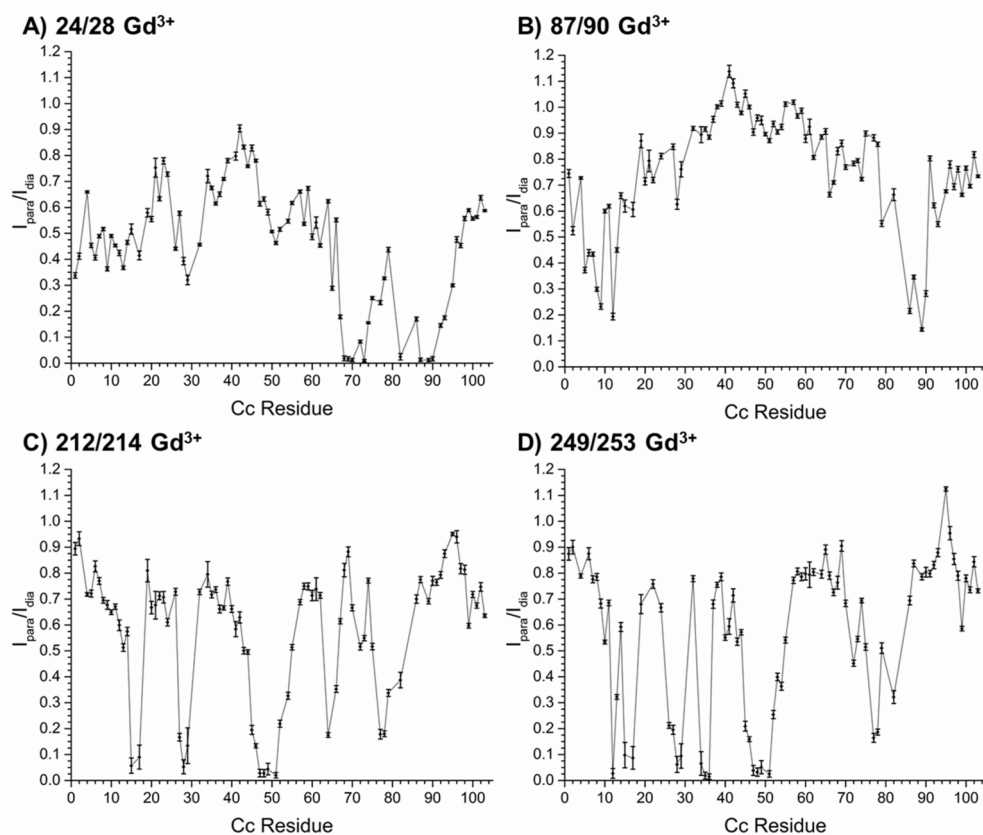


FIGURE 7.9 $I_{\text{para}}/I_{\text{dia}}$ values (black diamonds) for Cc amide protons in the presence of CcP with CLaNP-5-Gd³⁺ attached at position 24/28 (**A**), 87/90 (**B**), 212/214 (**C**) or 249/253 (**D**). The data were obtained with 78.4%, 72.7%, 73.5% or 76.5% CcP bound, respectively. The standard deviation of the noise level was determined for each spectrum and was then used in the calculation of the propagated error (black bars) for each $I_{\text{para}}/I_{\text{dia}}$ value.

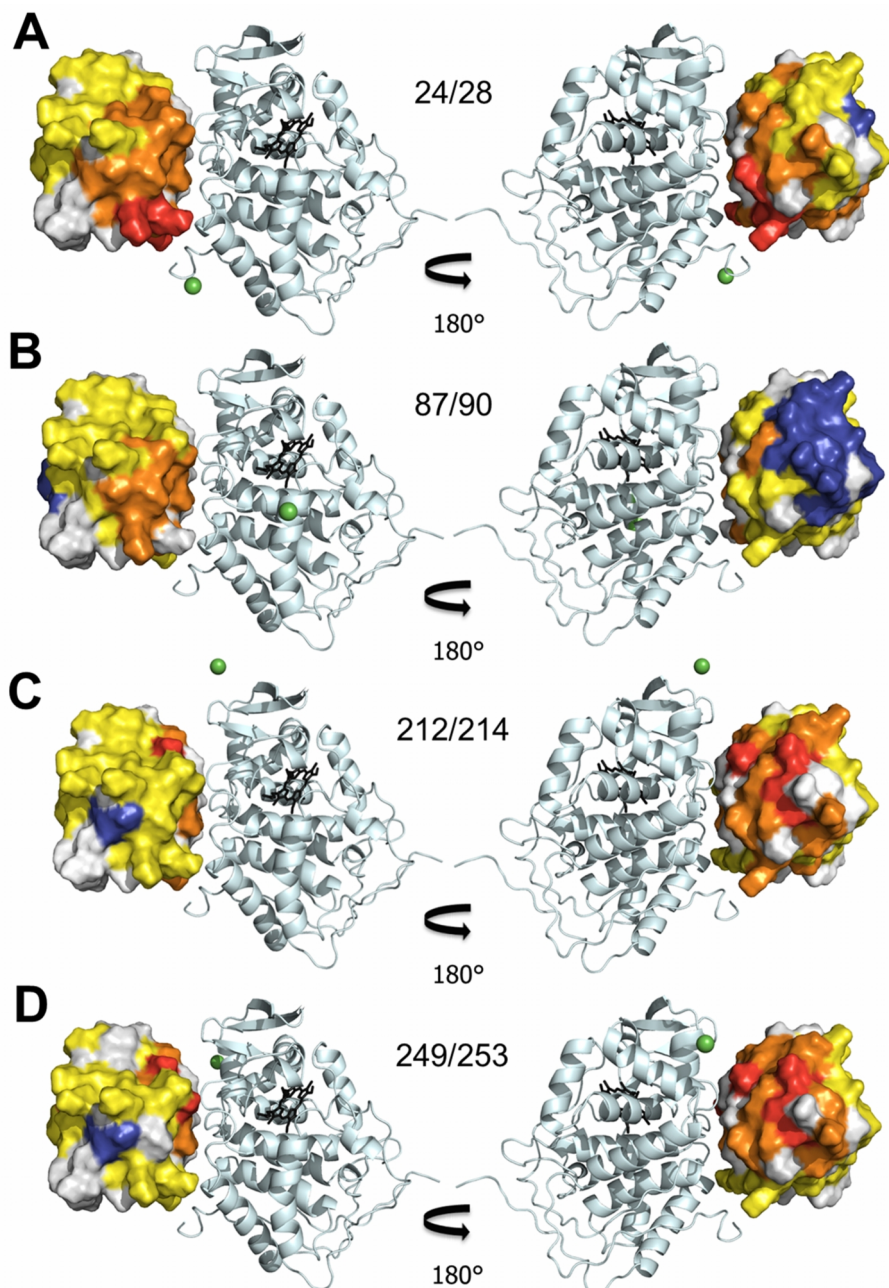


FIGURE 7.10 PRE map for Cc in a 1:1 complex with CLaNP-5-Gd³⁺ tagged CcP at positions 24/28 (**A**), 87/90 (**B**), 212/214 (**C**) or 249/253 (**D**). The PRE effects are colour-coded on a surface model of Cc in complex with CcP (grey ribbon; haem group in black sticks) (PDB-entry 2PCC)(5) and the Gd³⁺ ion in green. Residues with Γ_2 , para ≥ 200 s⁻¹ are red, 200 s⁻¹ < Γ_2 , para < 100 s⁻¹ are orange, 10 s⁻¹ < Γ_2 , para < 100 s⁻¹ are yellow, Γ_2 , para ≤ 10 s⁻¹ are blue and with no data are grey

Previous pNMR studies of the Cc-CcP complex used nitroxide based SLs to generate the PRE effects. These SLs only generate measurable effects between approximately 14-24 Å,⁽⁶²⁾ while the range for Gd³⁺ is approximately 20-45 Å, resulting in almost complete coverage of Cc. While the strongest effects are always located on the side of Cc facing the binding interface, which is consistent with the CSP map (Figure 7.3), the localization of the strongest effects (shown in red in both figures) has shifted to the residues nearest the CLaNP-5 probe.

The PRE effects were converted to distances between the affected residues and the paramagnetic centres and these were compared to back-predicted distances for three models of the complex (Figure 7.11). As discussed above, the PCS and RDC data contain additional orientation information that is clearly not described well by the 30:70 model of the complex.^(4, 8) Although the fits are not great, as discussed below, there is a better fit overall between the PRE derived distances and back predicted distances for the 30:70 model in comparison to the PCS and RDC data, which shows how insensitive PRE is to differences in orientation (Figures 7.7 & 7.8).

The best fit is seen for position 249/253 which fits the 30:70 model very well with the exception of residues 34-36 and 95-105 (Figure 7.11 D). The Tm³⁺ PCS and RDC data for position 249/253 also had a reasonably good fit to the back-predicted data for the 30:70 model, although the location of the strongest PCS were shifted slightly (Figure 7.7 G, H). Since limited averaging of the RDC was observed (Figure 7.8 D), these differences were likely due to changes in the stereo-specific complex caused by the presence of the CLaNP-5 probe. Similar results were seen for position 24/28. There was no significant averaging of the RDC (Figure 7.8 A) and the Tm³⁺ PCS fit the back-predicted data for the 30/70 model well; the fit appears to be less good for residues 65-80 and 85-95 but their predicted PCS were very large and therefore very few data points could be obtained for these nuclei (Figure 7.7 B). The PRE data also fit the 30:70 model well although several residues between 19-55 were much closer to the paramagnetic centre than predicted (Figure 7.11A). These results indicate that the probe is causing some minor disruptions to the stereo-specific complex.

In contrast, the data for position 87/90 suggest that the encounter state is being disrupted and its population is also likely increased. Both the PCS and RDC experience significant averaging (Figures 7.7 C, D and 7.8 B) and the PRE data show experimental distances for almost all nuclei that are less than predicted by any of the models (Figure 7.11 B). If the presence of the CLaNP-5 probe causes slight repulsion and rotation of Cc, these small differences in the distances to the paramagnetic probe will be measured by the extremely

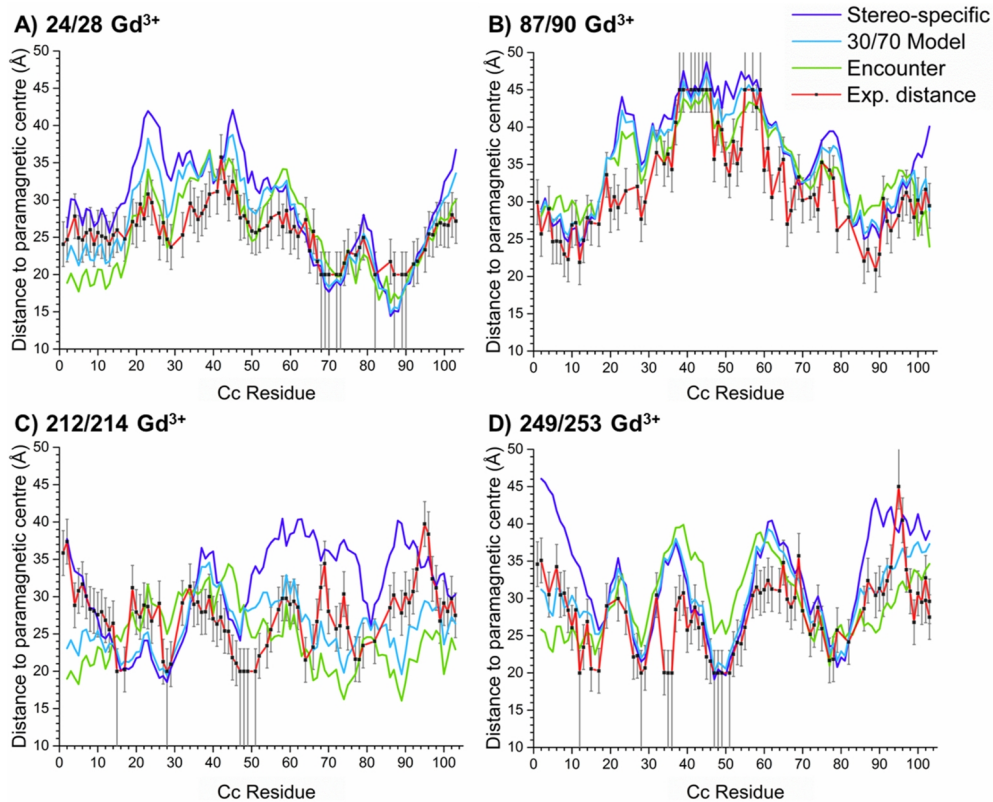


FIGURE 7.11 Analysis of intermolecular PRE data observed for Cc amide nitrogen atoms in a 1:1 complex with CLaNP-5-Ln³⁺ tagged CcP at positions 24/28 (**A, B**), 87/90 (**C, D**), 212/214 (**E, F**) or 249/253 (**G, H**). The experimental distances (black dots connected by red line) are compared to the theoretical distances for a 100% stereo-specific complex (purple line), a 100% encounter complex (green line) or a 30% encounter/70% stereo-specific complex (blue line).

sensitive PRE. The PRE map also shows a weaker interaction between the proteins, with Cc being pushed away from the CLaNP-5 attachment site (Figure 7.10 B) This likely indicates a greater proportion of the complex in the loosely associated and highly dynamic encounter state where Cc is sampling more of the surface of CcP than expected.

Similarly, for position 212/214 a significant degree of averaging of the PCS and RDC data as well as a very poor fit to the back-predicted data was observed (Figures 7.7 E, F and 7.8 C). The fit of the PRE data for this position was also poor; although the overall pattern of effects matches that of the 30:70 model relatively well, several regions (residues 1-13, 68-70, 93-97) are further away from the paramagnetic centre than predicted, matching more closely to the stereo-specific state, while others (residues 19-26, 48-55, 64, 77-82) are much closer than

predicted, matching more closely to the encounter state (Figure 7.11 C). Placing the probe at this position clearly interferes with complex formation and also appears to increase the population of the encounter complex.

In conclusion, the presence of the CLaNP-5 probe is disrupting complex formation to a greater or lesser degree for all double mutants. This is likely due to steric and electrostatic interference with Cc preventing the stereo-specific complex from forming, particularly for positions 24/28 and 249/253. The net positive charge of CLaNP-5 is not ideal when working with such highly charged proteins whose interaction is driven predominately by electrostatics.^(169, 328) For positions 87/90 and 212/214, the presence of the probe is also altering the encounter complex and likely increasing its population. Therefore, the existing models underestimate the spread of the Cc orientations sampled by this altered encounter state resulting in very poor fits to the PCS and RDC data, which are orientation sensitive. However, the PRE data fit the 30:70 model reasonably well, with some exceptions, demonstrating how insensitive PRE are to protein orientations, as we also concluded in Chapter 5.⁽³²¹⁾ Nevertheless, a new model that combines all three types of paramagnetic information will likely produce a much more accurate picture of how the Cc-CcP complex behaves in solution and so this work should be repeated with CLaNP-7 or other less charged probes.⁽³²⁹⁾ It would also be useful to combine the current Gd³⁺ generated PRE data with the previous nitroxide radical generated PRE data since the range in which these two paramagnetic centres generate effects are complementary: 20-45 Å and 14-24 Å, respectively. However, combining all of these effects cannot easily be done in the currently available protein docking software which presents a large technical challenge. Furthermore, although additional data sets are ideal for combating the *ill-posed inverse problem*,^(273, 274) this will require significant computational capacity.

CHAPTER 8

General conclusions & perspectives

GENERAL CONCLUSIONS & PERSPECTIVES

CYTOCHROME C PEROXIDASE

CcP has been the subject of scientific curiosity for almost eight decades since its discovery in 1940⁽¹³⁴⁾ but a backbone amide NMR resonance assignment was not possible until recently due to poor thermo-stability of the protein, especially at low ionic strengths.⁽²¹⁵⁾ In this work, we present a new expression construct and an updated purification protocol that produces very high yields of both natural abundance and isotope enriched protein that has enough thermo-stability to undergo several days of NMR measurements (Chapter 3).⁽¹⁴⁰⁾ This was necessary to achieve near-complete resonance back-bone assignment of CcP,⁽¹⁴⁰⁾ which was completed independently but concurrently with another group.⁽²¹⁶⁾ However, during these long term NMR experiments, a second minor form of the protein was detected in the NMR spectra, which seemed to result from sample aging (Chapter 3).

Like other haem proteins, the haem group in CcP can exist in several states depending on the nature of the axial ligand bound. These states occur during the steps in the enzymatic cycle but the affinity for different ligands can also be influenced by environmental conditions as well as the age of the enzyme (Chapter 2).⁽¹⁴⁶⁾ Using a combination of NMR and EPR, we demonstrated that a minor form observed in aged samples is a result of fluoride coordination to the haem resulting in a hexacoordinate, HS iron ion. The $\Delta\chi$ -tensor for fluoride bound CcP is similar to, but smaller in magnitude than, the $\Delta\chi$ -tensor for the RS enzyme. We show that the RS enzyme also contains a hexacoordinate, HS haem, contrary to the numerous reports stating that freshly prepared enzyme is pentacoordinate HS.^(146, 220) We are now the third group that has shown this using NMR.^(216, 225) For RS CcP, the $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values were similar to those published previously⁽²¹⁶⁾ at $-4.84 \pm 0.03 \times 10^{-32} \text{ m}^3$ and $-0.29 \pm 0.04 \times 10^{-32} \text{ m}^3$, respectively, while for fluoride bound CcP they were $-3.21 \pm 0.07 \times 10^{-32} \text{ m}^3$ and $-0.09 \pm 0.08 \times 10^{-32} \text{ m}^3$, respectively (Chapter 3).

The fluoride bound species does not occur in fresh samples and has only been seen in NMR samples that have been aged for several days at or above room temperature (Chapter 3). At pH 6, HF or F⁻ binds rapidly to the haem group⁽²³¹⁾ so the appearance of this minor species over time suggests that fluoride is being released slowly from another compound or solid material.⁽²⁷²⁾ Unfortunately, despite our efforts, we have not been able to identify the source of

the fluoride. This underscores the importance of using fresh samples for every experiment in order to ensure validity of the results.

Another consideration that must be taken into account when working with CcP at high concentrations ($\geq 400 \mu\text{M}$), such as those used during NMR experiments, is CcP exhibits ultra-weak self-association (Chapter 6). During experiments designed to measure intramolecular PRE (Chapter 5), many unexpected PRE were observed for residues more than 24 Å away from the SL. These effects could be diminished by the addition of Cc, confirming that they were indeed intermolecular PRE, and therefore the result of a CcP-CcP interaction. The PRE map shows that the interaction occurs between the stereo-specific binding interface on the “front” of CcP and several small, positively charged patches on the “back” of CcP, centred on residue K278 (Chapter 6). The spread of strong PREs indicates that the association represents an ensemble of orientations and not a single, well-defined CcP-CcP complex. Ultra-weak self-associations have been shown to occur with other proteins at similar concentrations (200-500 μM).^(304, 305, 318, 319) CcP has been thought to exist only as a monomer in solution, so the discovery of this ultra-weak self-association may have implications for future experiments.

SPIN LABEL MOBILITY

Paramagnetic SLs provide a convenient method for introducing site-specific relaxation centres to proteins, in the form of nitroxide radicals.⁽²⁸⁵⁾ However, the common methanethiosulfonate based SLs, such as MTSL, connect to proteins via a long, flexible linker resulting in a highly mobile tag that occupies several conformations in solution and therefore the unpaired electron is spatially distributed.⁽²⁸⁶⁾ Over- or underestimation of this distribution will decrease the accuracy of PRE results, but adequate modelling of the SL’s behaviour is difficult.⁽⁸⁹⁾ One solution is to limit tag dynamics by the addition of bulky side groups to the pyrroline ring as was done to create pyMTSL. A previous study⁽²⁹¹⁾ could fit PRE data generated by pyMTSL using a single conformer. We indeed found pyMTSL to be less mobile but, in our case, both MTSL and pyMTSL had to be modelled by an ensemble of conformers to produce the most accurate docking results (Chapter 5).

To help solve the problem of modelling the spin label distribution, the most favourable SL orientations can be experimentally determined using intra-molecular PRE data and then fixing the SL conformers in those positions during docking.^(28, 287-290) The SL

orientations were successfully determined for both MTSL and pyMTSL followed by protein docking based on the inter-molecular PRE data in which the conformers were either fixed in these orientations, fixed in random (sterically allowed) orientations or were allowed to rotate freely (Chapter 5). The best fits between the experimental and back-calculated intermolecular PRE were found when the SL orientations were free to rotate indicating that fixing the SL orientations and, thereby, limiting the degrees of freedom, cannot completely describe the observed data. However, assuming that the stereo-specific orientation of Cc observed in the crystal structure⁽¹⁶⁹⁾ is the dominant form in solution, the most accurate docking solutions were indeed found when the SLs were fixed in the experimentally determined most favourable orientations using intramolecular PRE (Chapter 5). Of course, this assumption is not entirely correct given that approximately 30% of the complex in solution exists in the encounter state,^(1, 8, 35, 140, 204) explaining some of the discrepancies in the fit.

Fixing the SL orientations during docking produces highly defined solutions as was seen previously.⁽¹⁾ However, precision is of no importance in the absence of accuracy, which can be difficult to achieve when the fixed SL orientations must be selected without experimental data, as is evident for the orientations that were fixed based only on steric information (Chapter 5). We showed that fixing the SLs in the experimentally determined most favourable orientations improved the accuracy of the final docking solutions showing that the results are highly dependent on the chosen SL orientations. SL flexibility appears to be the limiting factor for obtaining accurate PRE based docking results, at least in a dynamic protein complex such as the Cc-CcP complex.

This work demonstrated that PRE alone are not sufficient for accurate rigid body docking of a dynamic protein complex. Under most conditions, a large degree of rotation of Cc around the stereo-specific binding interface was observed in the final docking solutions. PRE are highly sensitive to minor states,⁽⁵⁴⁾ allowing lowly populated orientations to disproportionately influence the final results. This effect was lessened when the SLs were fixed in the experimentally determined most favourable orientations but overall PRE restraints are simply not for reliable for determining protein orientation within a complex (Chapter 5 & 7). Much better results can be obtained when PRE are combined with RDC or PCS data.^(14, 31, 80, 104-108, 299-301) Alternatively, good results can also be obtained when NOEs are combined with RDCs.^(80, 104-108, 299, 300, 302)

PARAMAGNETIC NMR AND ENCOUNTER COMPLEXES

Encounter complexes consist of several lowly populated conformations that cannot be isolated making them very difficult to study. NMR spectroscopy can be used to obtain structural information for these transient states but the observed data represent an average of all conformations present in solution, resulting in an ensemble of possible solutions instead of a precise description of the complex. This presents an inverse problem, in which many solutions can be found that fit the data. The only way to eliminate poor solutions is with additional data sets (Chapter 1). In this work, we have indeed shown that obtaining PRE data from “the other side” of the Cc-CcP complex results in novel effects that were not seen in previous studies. This highlights the importance of obtaining a comprehensive data set when working with encounter complexes (Chapter 4).

Diamagnetic NMR techniques such as CSP mapping can provide information on the degree of dynamics that occurs within a complex.^(9, 15, 27, 60, 170, 332) They can also be used to map the binding interface and calculate the binding or dissociation constants, as we have done here to assess the influence of SLs or CLaNP probes on the Cc-CcP interaction (Chapter 4 & 7). However, these techniques lack the sensitivity that pNMR provides, which is clearly demonstrated when comparing a CSP and a PRE map (Chapter 4). Furthermore, the CSP for weakly associated complexes can be small and difficult to discern.⁽³³⁶⁾ This is where pNMR, and particularly PRE, can play a vital role and they have been used to successfully characterize multiple ultra-weak complexes⁽³¹⁷⁻³¹⁹⁾ with K_D values as high as 25 mM.⁽³²⁰⁾ PRE was also used here to identify an ultra-weak self-association between CcP molecules, which until now have been assumed to exist as monomers in solution (Chapter 6). The existence of this CcP-CcP complex will need to be taken into account when designing future experiments, particularly when studying weak effects.

Alas, as discussed above, PRE are limited in their ability to accurately determine protein orientations within a complex during rigid body docking (Chapter 5). PRE are exquisitely sensitive to minor states but they only provide distance information. PCS provide both distance and orientation information. RDC provide only orientation information but the effect is not distance dependent and can therefore be very useful for visualizing a protein complex (Chapter 2). When used in combination, these effects can provide much more accurate structure determination than PRE alone.^(14, 31, 80, 104-108, 299-301) Conveniently, CLaNP probes containing different lanthanoids^(85, 102, 326) allow for the collection of PRE, PCS and RDC data using the same attachment site. This limits the need for time consuming

mutagenesis and with CLaNP-5- Ln^{3+} probes at four positions on the surface of CcP, we obtained essentially complete coverage of the Cc-CcP complex.

The previous model of the Cc-CcP complex in solution, in which the complex spent 70% of the time in the stereo-specific state observed in the crystal structure⁽¹⁶⁹⁾ and the remaining 30% in an encounter state, was fit using only PRE data.^(1, 8) In this work, we measured PRE data on CcP generated by SLs on Cc (Chapter 4) as well as PRE data on Cc generated by CLaNP-5- Gd^{3+} on CcP (Chapter 7). We found that both of these new data sets match the previous 30:70 model reasonably well. However, when the CLaNP-5 probe was used to generate orientation sensitive PCS and RDC on Cc, the resulting data were very poorly described by the 30:70 model (Chapter 7). Although this may indicate limitations in the ability of the 30:70 model to describe protein orientations within the complex, it can also be that the presence of a 3+ net charge on the CLaNP-5 probe caused electrostatic repulsion of the positively charged Cc and, therefore, significantly interfered with the formation of the complex. Such a highly charged probe is clearly not ideal when working with a complex whose interaction is driven predominately by electrostatics^(169, 328) and future studies should be done using CLaNP-7 or other less charged probes.⁽³²⁹⁾ Nevertheless, we recommend that, although technically challenging, a new model for the behaviour of the Cc-CcP complex in solution should be developed that incorporates all three types of paramagnetic information. This will hopefully lead to a more detailed description of the Cc-CcP encounter complex and perhaps also of the secondary Cc-CcP interaction via the low-affinity binding site.^(188-190, 200)

REFERENCES

- Volkov A.N., Worrall J.A., Holtzmann E., and Ubbink M., *Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR*. Proc Natl Acad Sci USA, 2006. **103**(50): p. 18945-18950.
- Ubbink M., *Dynamics in transient complexes of redox proteins*. Biochem Soc Trans, 2012. **40**(2): p. 415-418.
- Kim Y.C., Tang C., Clore G.M., and Hummer G., *Replica exchange simulations of transient encounter complexes in protein-protein association*. Proc Natl Acad Sci USA, 2008. **105**(35): p. 12855-12860.
- Selzer T. and Schreiber G., *New insights into the mechanism of protein-protein association*. Proteins, 2001. **45**(3): p. 190-198.
- Ubbink M., *The courtship of proteins: understanding the encounter complex*. FEBS Lett, 2009. **583**(7): p. 1060-1066.
- McLendon G., *Control of biological electron transport via molecular recognition and binding: the 'Velcro' model*. Struc Bonding, 1991. **75**: p. 159-174.
- Bashir Q., Scanu S., and Ubbink M., *Dynamics in electron transfer protein complexes*. FEBS J, 2011. **278**(9): p. 1391-1400.
- Bashir Q., Volkov A.N., Ullmann G.M., and Ubbink M., *Visualization of the encounter ensemble of the transient electron transfer complex of cytochrome c and cytochrome c peroxidase*. J Am Chem Soc, 2010. **132**(1): p. 241-247.
- Volkov A.N., Bashir Q., Worrall J.A., Ullmann G.M., and Ubbink M., *Shifting the equilibrium between the encounter state and the specific form of a protein complex by interfacial point mutations*. J Am Chem Soc, 2010. **132**(33): p. 11487-11495.
- Xiong P., Nocek J.M., Griffin A.K., Wang J., and Hoffman B.M., *Electrostatic redesign of the [myoglobin, cytochrome b₅] interface to create a well-defined docked complex with rapid interprotein electron transfer*. J Am Chem Soc, 2009. **131**(20): p. 6938-6939.
- Nocek J.M., Knutson A.K., Xiong P., Co N.P., and Hoffman B.M., *Photoinitiated singlet and triplet electron transfer across a redesigned [myoglobin, cytochrome b₅] interface*. J Am Chem Soc, 2010. **132**(17): p. 6165-6175.
- Xiong P., Nocek J.M., Vura-Weis J., Lockard J.V., Wasielewski M.R., and Hoffman B.M., *Faster interprotein electron transfer in a [myoglobin, b₅] complex with a redesigned interface*. Science, 2010. **330**(6007): p. 1075-1078.
- Hulsker R., Baranova M.V., Bullerjahn G.S., and Ubbink M., *Dynamics in the transient complex of plastocyanin-cytochrome f from Prochlorothrix hollandica*. J Am Chem Soc, 2008. **130**(6): p. 1985-1991.
- Xu X., Reinle W., Hannemann F., Konarev P.V., Svergun D.I., Bernhardt R., and Ubbink M., *Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy*. J Am Chem Soc, 2008. **130**(20): p. 6395-6403.
- Worrall J.A., Liu Y., Crowley P.B., Nocek J.M., Hoffman B.M., and Ubbink M., *Myoglobin and cytochrome b₅: a nuclear magnetic resonance study of a highly dynamic protein complex*. Biochemistry, 2002. **41**(39): p. 11721-11730.
- Liang Z.X., Jiang M., Ning Q., and Hoffman B.M., *Dynamic docking and electron transfer between myoglobin and cytochrome b₅*. J Biol Inorg Chem, 2002. **7**(6): p. 580-588.
- Liang Z.X., Nocek J.M., Huang K., Hayes R.T., Kurnikov I.V., Beratan D.N., and Hoffman B.M., *Dynamic docking and electron*

- transfer between Zn-myoglobin and cytochrome b₅*. J Am Chem Soc, 2002. **124**(24): p. 6849-6859.
18. Liang Z.X., Nocek J.M., Kurnikov I.V., Beratan D.N., and Hoffman B.M., *Electrostatic control of electron transfer between myoglobin and cytochrome b₅: Effect of methylating the heme propionates of Zn-myoglobin*. J Am Chem Soc, 2000. **122**(14): p. 3552-3553.
 19. Schlarb-Ridley B.G., Bendall D.S., and Howe C.J., *Role of electrostatics in the interaction between cytochrome f and plastocyanin of the cyanobacterium Phormidium laminosum*. Biochemistry, 2002. **41**(10): p. 3279-3285.
 20. Hart S.E., Schlarb-Ridley B.G., Delon C., Bendall D.S., and Howe C.J., *Role of charges on cytochrome f from the cyanobacterium Phormidium laminosum in its interaction with plastocyanin*. Biochemistry, 2003. **42**(17): p. 4829-4836.
 21. Schlarb-Ridley B.G., Mi H.L., Teale W.D., Meyer V.S., Howe C.J., and Bendall D.S., *Implications of the effects of viscosity, macromolecular crowding, and temperature for the transient interaction between cytochrome f and plastocyanin from the cyanobacterium Phormidium laminosum*. Biochemistry, 2005. **44**(16): p. 6232-6238.
 22. Schlarb-Ridley B.G., Bendall D.S., and Howe C.J., *Relation between interface properties and kinetics of electron transfer in the interaction of cytochrome f and plastocyanin from plants and the cyanobacterium Phormidium laminosum*. Biochemistry, 2003. **42**(14): p. 4057-4063.
 23. Crowley P.B., Otting G., Schlarb-Ridley B.G., Canters G.W., and Ubbink M., *Hydrophobic interactions in a cyanobacterial plastocyanin-cytochrome f complex*. J Am Chem Soc, 2001. **123**(43): p. 10444-10453.
 24. Nocek J.M., Sishta B.P., Cameron J.C., Mauk A.G., and Hoffmann B.M., *Cyclic electron transfer within the [Zn-myoglobin, cytochrome b₅] complex*. J Am Chem Soc, 1997. **119**(9): p. 2146-2155.
 25. Harel M., Cohen M., and Schreiber G., *On the dynamic nature of the transition state for protein-protein association as determined by double-mutant cycle analysis and simulation*. J Mol Biol, 2007. **371**(1): p. 180-196.
 26. Ubbink M. and Bendall D.S., *Complex of plastocyanin and cytochrome c characterized by NMR chemical shift analysis*. Biochemistry, 1997. **36**(21): p. 6326-6335.
 27. Worrall J.A., Reinle W., Bernhardt R., and Ubbink M., *Transient protein interactions studied by NMR spectroscopy: the case of cytochrome c and adrenodoxin*. Biochemistry, 2003. **42**(23): p. 7068-7076.
 28. Tang C., Iwahara J., and Clore G.M., *Visualization of transient encounter complexes in protein-protein association*. Nature, 2006. **444**(7117): p. 383-386.
 29. Iwahara J. and Clore G.M., *Detecting transient intermediates in macromolecular binding by paramagnetic NMR*. Nature, 2006. **440**(7088): p. 1227-1230.
 30. Tolman J.R., Flanagan J.M., Kennedy M.A., and Prestegard J.H., *NMR evidence for slow collective motions in cyanometmyoglobin*. Nat Struct Biol, 1997. **4**(4): p. 292-297.
 31. Xu X., Keizers P.H., Reinle W., Hannemann F., Bernhardt R., and Ubbink M., *Intermolecular dynamics studied by paramagnetic tagging*. J Biomol NMR, 2009. **43**(4): p. 247-254.
 32. Bertini I., Giachetti A., Luchinat C., Parigi G., Petoukhov M.V., Pierattelli R., Ravera E., and Svergun D.I., *Conformational space of flexible biological macromolecules from average data*. J Am Chem Soc, 2010. **132**(38): p. 13553-13558.
 33. Dasgupta S., Hu X., Keizers P.H., Liu W.M., Luchinat C., Nagulapalli M., Overhand M., Parigi G., Sgheri L., and Ubbink M., *Narrowing the conformational space sampled by two-domain proteins with paramagnetic*

References

- probes in both domains.* J Biomol NMR, 2011. **51**(3): p. 253-263.
34. Fawzi N.L., Doucleff M., Suh J.Y., and Clore G.M., *Mechanistic details of a protein-protein association pathway revealed by paramagnetic relaxation enhancement titration measurements.* Proc Natl Acad Sci USA, 2010. **107**(4): p. 1379-1384.
 35. Volkov A.N., Ubbink M., and van Nuland N.A., *Mapping the encounter state of a transient protein complex by PRE NMR spectroscopy.* J Biomol NMR, 2010. **48**(4): p. 225-236.
 36. Miyashita O., Onuchic J.N., and Okamura M.Y., *Transition state and encounter complex for fast association of cytochrome *c*₂ with bacterial reaction center.* Proc Natl Acad Sci USA, 2004. **101**(46): p. 16174-16179.
 37. Schreiber G., Haran G., and Zhou H.X., *Fundamental aspects of protein-protein association kinetics.* Chem Rev, 2009. **109**(3): p. 839-860.
 38. Northrup S.H., Boles J.O., and Reynolds J.C., *Brownian dynamics of cytochrome *c* and cytochrome *c* peroxidase association.* Science, 1988. **241**(4861): p. 67-70.
 39. Camacho C.J., Weng Z., Vajda S., and DeLisi C., *Free energy landscapes of encounter complexes in protein-protein association.* Biophys J, 1999. **76**(3): p. 1166-1178.
 40. Camacho C.J., Kimura S.R., DeLisi C., and Vajda S., *Kinetics of desolvation-mediated protein-protein binding.* Biophys J, 2000. **78**(3): p. 1094-1105.
 41. Camacho C.J. and Vajda S., *Protein docking along smooth association pathways.* Proc Natl Acad Sci USA, 2001. **98**(19): p. 10636-10641.
 42. Spaar A. and Helms V., *Free energy landscape of protein-protein encounter resulting from Brownian dynamics simulations of Barnase:Barstar.* J Chem Theory Comput, 2005. **1**(4): p. 723-736.
 43. Spaar A., Dammer C., Gabdoulline R.R., Wade R.C., and Helms V., *Diffusional encounter of barnase and barstar.* Biophys J, 2006. **90**(6): p. 1913-1924.
 44. Volkov A.N. and van Nuland N.A.J., *Electron transfer interactions of cytochrome *c*.* PLoS Comput Biol, 2012. **8**(12).
 45. Scanu S., Foerster J.M., Ullmann G.M., and Ubbink M., *Role of hydrophobic interactions in the encounter complex formation of the plastocyanin and cytochrome *f* complex revealed by paramagnetic NMR spectroscopy.* J Am Chem Soc, 2013. **135**(20): p. 7681-7692.
 46. Liang Z.X., Kurnikov I.V., Nocek J.M., Mauk A.G., Beratan D.N., and Hoffman B.M., *Dynamic docking and electron transfer between cytochrome *b*₅ and a suite of myoglobin surface-charge mutants. Introduction of a functional-docking algorithm for protein-protein complexes.* J Am Chem Soc, 2004. **126**(9): p. 2785-2798.
 47. Trana E.N., Nocek J.M., Knutson A.K., and Hoffman B.M., *Evolving the [myoglobin, cytochrome *b*₅] complex from dynamic toward simple docking: charging the electron transfer reactive patch.* Biochemistry, 2012. **51**(43): p. 8542-8553.
 48. Sugase K., Dyson H.J., and Wright P.E., *Mechanism of coupled folding and binding of an intrinsically disordered protein.* Nature, 2007. **447**(7147): p. 1021-1025.
 49. Ubbink M., Ejdeback M., Karlsson B.G., and Bendall D.S., *The structure of the complex of plastocyanin and cytochrome *f*, determined by paramagnetic NMR and restrained rigid-body molecular dynamics.* Structure, 1998. **6**(3): p. 323-335.
 50. Ejdeback M., Bergkvist A., Karlsson B.G., and Ubbink M., *Side-chain interactions in the plastocyanin-cytochrome *f* complex.* Biochemistry, 2000. **39**(17): p. 5022-5027.
 51. Diaz-Moreno I., Diaz-Quintana A., De la Rosa M.A., and Ubbink M., *Structure of the complex between plastocyanin and cytochrome *f* from the cyanobacterium *Nostoc sp* PCC 7119 as*

- determined by paramagnetic NMR - The balance between electrostatic and hydrophobic interactions within the transient complex determines the relative orientation of the two proteins. *J Biol Chem*, 2005. **280**(19): p. 18908-18915.
52. Albarran C., Navarro J.A., Molina-Heredia F.P., Murdoch P.S., De la Rosa M.A., and Hervas M., *Laser flash-induced kinetic analysis of cytochrome f oxidation by wild-type and mutant plastocyanin from the cyanobacterium Nostoc sp PCC 7119*. *Biochemistry*, 2005. **44**(34): p. 11601-11607.
 53. Albarran C., Navarro J.A., De la Rosa M.A., and Hervas M., *The specificity in the interaction between cytochrome f and plastocyanin from the cyanobacterium Nostoc sp PCC 7119 is mainly determined by the copper protein*. *Biochemistry*, 2007. **46**(4): p. 997-1003.
 54. Scanu S., Forster J., Finiguerra M.G., Shabestari M.H., Huber M., and Ubbink M., *The complex of cytochrome f and plastocyanin from Nostoc sp PCC 7119 is highly dynamic*. *Chembiochem*, 2012. **13**(9): p. 1312-1318.
 55. Baniulis D., Yamashita E., Whitelegge J.P., Zatsman A.I., Hendrich M.P., Hasan S.S., Ryan C.M., and Cramer W.A., *Structure-function, stability, and chemical modification of the cyanobacterial cytochrome b₆f complex from Nostoc sp. PCC 7120*. *J Biol Chem*, 2009. **284**(15): p. 9861-9869.
 56. Harel M., Spaar A., and Schreiber G., *Fruitful and futile encounters along the association reaction between proteins*. *Biophys J*, 2009. **96**(10): p. 4237-4248.
 57. Schilder J. and Ubbink M., *Formation of transient protein complexes*. *Curr Opin Struct Biol*, 2013. **23**(6): p. 911-918.
 58. Schilder J., Hass M.A.S., Keizers P.H., and Ubbink M., *Paramagnetic NMR spectroscopy and lowly populated states*. (M. G. Clore & J. Potts, Eds.). *Recent Developments in Biomolecular NMR 2012*: Royal Society of Chemistry. 130-150.
 59. Clore G.M., *Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation*. *Protein Sci*, 2011. **20**(2): p. 229-246.
 60. Volkov A.N., Bashir Q., Worrall J.A.R., and Ubbink M., *Binding hot spot in the weak protein complex of physiological redox partners yeast cytochrome c and cytochrome c peroxidase*. *J Mol Biol*, 2009. **385**(3): p. 1003-1013.
 61. Otting G., *Protein NMR using paramagnetic ions*. *Annu Rev Biophys*, 2010. **39**: p. 387-405.
 62. Keizers P.H. and Ubbink M., *Paramagnetic tools in protein NMR*, in *Protein NMR spectroscopy: Practical techniques and applications* 2011, John Wiley & Sons, Ltd. p. 193-219.
 63. Bertini I., Luchinat C., and Aime S., *NMR of paramagnetic substances*. *Coordination Chemistry Reviews*, 1996. **150**: p. 29-110.
 64. Liu Y.Z., Zhang X.H., Yoshida T., and La Mar G.N., *Solution ¹H NMR characterization of the distal H-bond network and the effective axial field in the resting-state, high-spin ferric, substrate-bound complex of heme oxygenase from N. meningitidis*. *J Am Chem Soc*, 2005. **127**(17): p. 6409-6422.
 65. Schejter A., Lanir A., Vig I., and Cohen J.S., *Direct observation of the methionine residues of cytochrome c by ¹³C nuclear magnetic resonance spectroscopy*. *J Biol Chem*, 1978. **253**(11): p. 3768-3770.
 66. Grimaldi J.J. and Sykes B.D., *Concanavalin A: a stopped flow nuclear magnetic resonance study of conformational changes induced by Mn⁺⁺, Ca⁺⁺, and α-methyl-D-mannoside*. *J Biol Chem*, 1975. **250**(5): p. 1618-1624.
 67. Tu K. and Gochin M., *Structure determination by restrained molecular dynamics using NMR pseudocontact shifts as experimentally determined constraints*. *J Am Chem Soc*, 1999. **121**(40): p. 9276-9285.

References

68. Salgado J., Jimenez H.R., Moratal J.M., Kroes S., Warmerdam G.C., and Canters G.W., *Paramagnetic cobalt and nickel derivatives of *Alcaligenes denitrificans* azurin and its M121Q mutant. A ¹H NMR study.* *Biochemistry*, 1996. **35**(6): p. 1810-1819.
69. Beattie J.K., Fensom D.J., Freeman H.C., Woodcock E., Hill H.A., and Stokes A.M., *An NMR investigation of electron transfer in the copper-protein, plastocyanin.* *Biochim Biophys Acta*, 1975. **405**(1): p. 109-114.
70. Vlasie M.D., Comuzzi C., van den Nieuwendijk A.M., Prudencio M., Overhand M., and Ubbink M., *Long-range-distance NMR effects in a protein labeled with a lanthanide-DOTA chelate.* *Chemistry*, 2007. **13**(6): p. 1715-1723.
71. Pintacuda G., Park A.Y., Keniry M.A., Dixon N.E., and Otting G., *Lanthanide labeling offers fast NMR approach to 3D structure determinations of protein-protein complexes.* *J Am Chem Soc*, 2006. **128**(11): p. 3696-3702.
72. Horrocks W.D., Jr. and Sipe J.P., 3rd, *Lanthanide complexes as nuclear magnetic resonance structural probes: paramagnetic anisotropy of shift reagent adducts.* *Science*, 1972. **177**(4053): p. 994-996.
73. Prudencio M., Rohovec J., Peters J.A., Tocheva E., Boulanger M.J., Murphy M.E., Hupkes H.J., Kusters W., Impagliazzo A., and Ubbink M., *A caged lanthanide complex as a paramagnetic shift agent for protein NMR.* *Chemistry*, 2004. **10**(13): p. 3252-3260.
74. Su X.C., Huber T., Dixon N.E., and Otting G., *Site-specific labelling of proteins with a rigid lanthanide-binding tag.* *Chembiochem*, 2006. **7**(10): p. 1599-1604.
75. Feeny J., Birdsall B., Bradbury A.F., Biekofsky R.R., and Bayley P.M., *Calmodulin tagging provides a general method of using lanthanide induced magnetic field orientation to observe residual dipolar couplings in proteins in solution.* *J Biomol NMR*, 2001. **21**(1): p. 41-48.
76. Wohnert J., Franz K.J., Nitz M., Imperiali B., and Schwalbe H., *Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings.* *J Am Chem Soc*, 2003. **125**(44): p. 13338-13339.
77. Battiste J.L. and Wagner G., *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*, 2000. **39**(18): p. 5355-5365.
78. Cotton S., *Introduction to the lanthanides*, in *Lanthanide and actinide chemistry* 2006, John Wiley & Sons, Ltd. p. 1-7.
79. Madl T., Guttler T., Gorlich D., and Sattler M., *Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements.* *Angew Chem Int Edit*, 2011. **50**(17): p. 3993-3997.
80. Simon B., Madl T., Mackereth C.D., Nilges M., and Sattler M., *An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution.* *Angew Chem Int Edit*, 2010. **49**(11): p. 1967-1970.
81. Su X.C., Liang H.B., Loscha K.V., and Otting G., *[Ln(DPA)₃]³⁻ is a convenient paramagnetic shift reagent for protein NMR studies.* *J Am Chem Soc*, 2009. **131**(30): p. 10352-10353.
82. Yagi H., Loscha K.V., Su X.C., Stanton-Cook M., Huber T., and Otting G., *Tunable paramagnetic relaxation enhancements by [Gd(DPA)₃]³⁻ for protein structure analysis.* *J Biomol NMR*, 2010. **47**(2): p. 143-153.
83. Su X.C. and Otting G., *Paramagnetic labelling of proteins and oligonucleotides for NMR.* *J Biomol NMR*, 2010. **46**(1): p. 101-112.
84. Hass M.A.S., Keizers P.H.J., Blok A., Hiruma Y., and Ubbink M., *Validation of a lanthanide tag for the analysis of protein dynamics by paramagnetic NMR spectroscopy.* *J Am Chem Soc*, 2010. **132**(29): p. 9952-9953.

85. Keizers P.H., Desreux J.F., Overhand M., and Ubbink M., *Increased paramagnetic effect of a lanthanide protein probe by two-point attachment.* J Am Chem Soc, 2007. **129**(30): p. 9292-9293.
86. Yu D., Volkov A.N., and Tang C., *Characterizing dynamic protein-protein interactions using differentially scaled paramagnetic relaxation enhancement.* J Am Chem Soc, 2009. **131**(47): p. 17291-17297.
87. Solomon I., *Relaxation processes in a system of two spins.* Phys Rev, 1955. **99**(2): p. 559-565.
88. Gueron M., *Nuclear-relaxation in macromolecules by paramagnetic ions: A novel mechanism.* J Magn Reson, 1975. **19**(1): p. 58-66.
89. Iwahara J., Schwieters C.D., and Clore G.M., *Ensemble approach for NMR structure refinement against ¹H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule.* J Am Chem Soc, 2004. **126**(18): p. 5879-5896.
90. Iwahara J., Tang C., and Clore G.M., *Practical aspects of ¹H transverse paramagnetic relaxation enhancement measurements on macromolecules.* J Magn Reson, 2007. **184**(2): p. 185-195.
91. Geraldes C.F., *Lanthanide shift reagents.* Methods enzymol, 1993. **227**: p. 43-78.
92. Phillips W.D., Poe M., McDonald C.C., and Bartsch R.G., *Proton magnetic resonance studies of chromatium high-potential iron protein.* Proc Natl Acad Sci USA, 1970. **67**(2): p. 682-687.
93. Hass M.A.S. and Ubbink M., *Structure determination of protein-protein complexes with long-range anisotropic paramagnetic NMR restraints.* Curr Opin Cell Biol, 2014. **24**: p. 45-53.
94. Hiruma Y., Hass M.A.S., Kikui Y., Liu W.M., Olmez B., Skinner S.P., Blok A., Kloosterman A., Koteishi H., Lohr F., Schwalbe H., Nojiri M., and Ubbink M., *The structure of the cytochrome P450cam-putidaredoxin complex determined by paramagnetic NMR spectroscopy and crystallography.* J Mol Biol, 2013. **425**(22): p. 4353-4365.
95. Wang X., Srisailam S., Yee A.A., Lemak A., Arrowsmith C., Prestegard J.H., and Tian F., *Domain-domain motions in proteins from time-modulated pseudocontact shifts.* J Biomol NMR, 2007. **39**(1): p. 53-61.
96. Schmitz C., Stanton-Cook M.J., Su X.C., Otting G., and Huber T., *Nubat: an interactive software tool for fitting $\Delta\chi$ -tensors to molecular coordinates using pseudocontact shifts.* J Biomol NMR, 2008. **41**(3): p. 179-189.
97. Pintacuda G., Keniry M.A., Huber T., Park A.Y., Dixon N.E., and Otting G., *Fast structure-based assignment of ¹⁵N HSQC spectra of selectively ¹⁵N-labeled paramagnetic proteins.* J Am Chem Soc, 2004. **126**(9): p. 2963-2970.
98. Bax A., *Weak alignment offers new NMR opportunities to study protein structure and dynamics.* Protein Sci, 2003. **12**(1): p. 1-16.
99. Tjandra N., Omichinski J.G., Gronenborn A.M., Clore G.M., and Bax A., *Use of dipolar ¹H-¹⁵N and ¹H-¹³C couplings in the structure determination of magnetically oriented macromolecules in solution.* Nat Struct Biol, 1997. **4**(9): p. 732-738.
100. Man B., Su X.C., Liang H., Simonsen S., Huber T., Messerle B.A., and Otting G., *3-Mercapto-2,6-pyridinedicarboxylic acid: a small lanthanide-binding tag for protein studies by NMR spectroscopy.* Chemistry, 2010. **16**(12): p. 3827-3832.
101. Bertini I., Kursula P., Luchinat C., Parigi G., Vahokoski J., Wilmanns M., and Yuan J., *Accurate solution structures of proteins from X-ray data and a minimal set of NMR data: calmodulin-peptide complexes as examples.* J Am Chem Soc, 2009. **131**(14): p. 5134-5144.
102. Keizers P.H., Saragliadis A., Hiruma Y., Overhand M., and Ubbink M., *Design, synthesis, and evaluation of a lanthanide chelating*

References

- protein probe: CLaNP-5 yields predictable paramagnetic effects independent of environment.* J Am Chem Soc, 2008. **130**(44): p. 14802-14812.
103. Otting G., *Prospects for lanthanides in structural biology by NMR.* J Biomol NMR, 2008. **42**(1): p. 1-9.
 104. Salmon L., Nodet G., Ozenne V., Yin G.W., Jensen M.R., Zweckstetter M., and Blackledge M., *NMR characterization of long-range order in intrinsically disordered proteins.* J Am Chem Soc, 2010. **132**(24): p. 8407-8418.
 105. Bertoncini C.W., Jung Y.S., Fernandez C.O., Hoyer W., Griesinger C., Jovin T.M., and Zweckstetter M., *Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein.* Proc Natl Acad Sci USA, 2005. **102**(5): p. 1430-1435.
 106. Song J., Guo L.W., Muradov H., Artemyev N.O., Ruoho A.E., and Markley J.L., *Intrinsically disordered gamma-subunit of cGMP phosphodiesterase encodes functionally relevant transient secondary and tertiary structure.* Proc Natl Acad Sci USA, 2008. **105**(5): p. 1505-1510.
 107. Huang J.R. and Grzesiek S., *Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin.* J Am Chem Soc, 2010. **132**(2): p. 694-705.
 108. Leeper T.C., Qu X., Lu C., Moore C., and Varani G., *Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1.* J Mol Biol, 2010. **401**(3): p. 334-349.
 109. Bertini I., Del Bianco C., Gelis I., Katsaros N., Luchinat C., Parigi G., Peana M., Provenzani A., and Zoroddu M.A., *Experimentally exploring the conformational space sampled by domain reorientation in calmodulin.* Proc Natl Acad Sci USA, 2004. **101**(18): p. 6841-6846.
 110. Biekofsky R.R., Muskett F.W., Schmidt J.M., Martin S.R., Browne J.P., Bayley P.M., and Feeney J., *NMR approaches for monitoring domain orientations in calcium-binding proteins in solution using partial replacement of Ca²⁺ by Tb³⁺.* FEBS Lett, 1999. **460**(3): p. 519-526.
 111. Assfalg M., Bertini I., Turano P., Mauk A.G., Winkler J.R., and Gray H.B., *¹⁵N-¹H residual dipolar coupling analysis of native and alkaline-K79A Saccharomyces cerevisiae cytochrome c.* Biophys J, 2003. **84**(6): p. 3917-3923.
 112. Bertini I., Faraone-Mennella J., Gray H.B., Luchinat C., Parigi G., and Winkler J.R., *NMR-validated structural model for oxidized Rhodospirillum rubrum cytochrome c556.* J Biol Inorg Chem, 2004. **9**(2): p. 224-230.
 113. Keilin D., *On cytochrome, a respiratory pigment, common to animals, yeast, and higher plants.* Proc R Soc London B, 1925. **98**(690): p. 312-339.
 114. Keilin D., *Cytochrome and respiratory enzymes.* Proc R Soc London B, 1929. **104**(730): p. 206-252.
 115. Hartree E.F., *The discovery of cytochrome.* Biochem Educ, 1973. **1**(4): p. 69-71.
 116. Banci L. and Assfalg M., *Mitochondrial cytochrome c*, in *Handbook of Metalloproteins* 2006, John Wiley & Sons, Ltd.
 117. The UniProt Consortium, *Activities at the Universal Protein Resource (UniProt).* Nucleic Acids Res, 2014. **42**(D1): p. D191-D198.
 118. Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Soding J., Thompson J.D., and Higgins D.G., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.
 119. Sievers F. and Higgins D.G., *Clustal Omega, accurate alignment of very large numbers of sequences.* Methods Mol Biol, 2014. **1079**: p. 105-116.

120. Volkov A.N., Nicholls P., and Worrall J.A.R., *The complex of cytochrome c and cytochrome c peroxidase: The end of the road?* BBA-Bioenergetics, 2011. **1807**(11): p. 1482-1503.
121. Sherman F., Taber H., and Campbell W., *Genetic determination of iso-cytochromes c in yeast.* J Mol Biol, 1965. **13**(1): p. 21-39.
122. Sherman F., Stewart J.W., Parker J.H., Inhaber E., Shipman N.A., Putterman G.J., Gardisky R.L., and Margoliash E., *The mutational alteration of the primary structure of yeast iso-1-cytochrome c.* J Biol Chem, 1968. **243**(20): p. 5446-5456.
123. Willie A., McLean M., Liu R.Q., Hilgen-Willis S., Saunders A.J., Pielak G.J., Sligar S.G., Durham B., and Millett F., *Intracomplex electron transfer between ruthenium-65-cytochrome b₅ and position-82 variants of yeast iso-1-cytochrome c.* Biochemistry, 1993. **32**(29): p. 7519-7525.
124. Pollock W.B., Rosell F.I., Twitchett M.B., Dumont M.E., and Mauk A.G., *Bacterial expression of a mitochondrial cytochrome c. Trimethylation of lys72 in yeast iso-1-cytochrome c and the alkaline conformational transition.* Biochemistry, 1998. **37**(17): p. 6124-6131.
125. Morar A.S., Kakouras D., Young G.B., Boyd J., and Pielak G.J., *Expression of ¹⁵N-labeled eukaryotic cytochrome c in Escherichia coli.* J Biol Inorg Chem, 1999. **4**(2): p. 220-222.
126. Dumont M.E., Ernst J.F., Hampsey D.M., and Sherman F., *Identification and sequence of the gene encoding cytochrome c heme lyase in the yeast Saccharomyces cerevisiae.* EMBO J, 1987. **6**(1): p. 235-241.
127. Berghuis A.M. and Brayer G.D., *Oxidation state-dependent conformational changes in cytochrome c.* J Mol Biol, 1992. **223**(4): p. 959-976.
128. Louie G.V. and Brayer G.D., *High-resolution refinement of yeast iso-1-cytochrome c and comparisons with other eukaryotic cytochromes c.* J Mol Biol, 1990. **214**(2): p. 527-555.
129. Baistrocchi P., Banci L., Bertini I., Turano P., Bren K.L., and Gray H.B., *Three-dimensional solution structure of Saccharomyces cerevisiae reduced iso-1-cytochrome c.* Biochemistry, 1996. **35**(43): p. 13788-13796.
130. Louie G.V., Hutcheon W.L., and Brayer G.D., *Yeast iso-1-cytochrome c. A 2.8 Å resolution three-dimensional structure determination.* J Mol Biol, 1988. **199**(2): p. 295-314.
131. Banci L., Bertini I., Bren K.L., Gray H.B., Sompornpisut P., and Turano P., *Solution structure of oxidized Saccharomyces cerevisiae iso-1-cytochrome c.* Biochemistry, 1997. **36**(29): p. 8992-9001.
132. Bashir Q., *Dynamics in electron transfer protein complexes in Doctoral thesis*, 2010, Leiden University.
133. Liu X., Kim C.N., Yang J., Jemmerson R., and Wang X., *Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c.* Cell, 1996. **86**(1): p. 147-157.
134. Altschul A.M., Abrams R., and Hogness T.R., *Cytochrome c peroxidase.* J Biol Chem, 1940. **136**: p. 777-794.
135. Yonetani T., *Studies on cytochrome c peroxidase. II. Stoichiometry between enzyme, H₂O₂, and ferrocytochrome c and enzymic determination of extinction coefficients of cytochrome c.* J Biol Chem, 1965. **240**(11): p. 4509-4514.
136. Goodin D.B., Mauk A.G., and Smith M., *Studies of the radical species in compound ES of cytochrome c peroxidase altered by site-directed mutagenesis.* Proc Natl Acad Sci USA, 1986. **83**(5): p. 1295-1299.
137. Goodin D.B., Mauk A.G., and Smith M., *The peroxide complex of yeast cytochrome c peroxidase contains two distinct radical species, neither of which resides at methionine 172 or tryptophan 51.* J Biol Chem, 1987. **262**(16): p. 7719-7724.
138. Goltz S., Kaput J., and Blobel G., *Isolation of the yeast nuclear gene encoding the mitochondrial*

References

- protein, cytochrome c peroxidase.* J Biol Chem, 1982. **257**(18): p. 11186-11190.
139. Goodin D.B., Davidson M.G., Roe J.A., Mauk A.G., and Smith M., *Amino-acid substitutions at tryptophan-51 of cytochrome c peroxidase - effects on coordination, species preference for cytochrome c, and electron-transfer.* Biochemistry, 1991. **30**(20): p. 4953-4962.
140. Schilder J., Lohr F., Schwalbe H., and Ubbink M., *The cytochrome c peroxidase and cytochrome c encounter complex: the other side of the story.* Febs Letters, 2014. **588**(10): p. 1873-1878.
141. Volkov A.N., Wohlkonig A., Soror S.H., and van Nuland N.A.J., *Expression, purification, characterization, and solution nuclear magnetic resonance study of highly deuterated yeast cytochrome c peroxidase with enhanced solubility.* Biochemistry, 2013. **52**(13): p. 2165-2175.
142. Teske J.G., Savenkova M.I., Mauro J.M., Erman J.E., and Satterlee J.D., *Yeast cytochrome c peroxidase expression in Escherichia coli and rapid isolation of various highly pure holoenzymes.* Protein Expr Purif, 2000. **19**(1): p. 139-147.
143. Savenkova M.I., Satterlee J.D., Erman J.E., Siems W.F., and Helms G.L., *Expression, purification, characterization, and NMR studies of highly deuterated recombinant cytochrome c peroxidase.* Biochemistry, 2001. **40**(40): p. 12123-12131.
144. Bonagura C.A., Bhaskar B., Shimizu H., Li H.Y., Sundaramoorthy M., McRee D.E., Goodin D.B., and Poulos T.L., *High-resolution crystal structures and spectroscopy of native and compound I cytochrome c peroxidase.* Biochemistry, 2003. **42**(19): p. 5600-5608.
145. Iizuka T., Kotani M., and Yonetani T., *Magnetic susceptibility measurements of cytochrome c peroxidase and its complexes.* J Biol Chem, 1971. **246**(15): p. 4731-4736.
146. Yonetani T. and Anni H., *Yeast cytochrome c peroxidase. Coordination and spin states of heme prosthetic group.* J Biol Chem, 1987. **262**(20): p. 9547-9554.
147. Finzel B.C., Poulos T.L., and Kraut J., *Crystal structure of yeast cytochrome c peroxidase refined at 1.7 Å resolution.* J Biol Chem, 1984. **259**(21): p. 13027-13036.
148. Bonagura C.A., Bhaskar B., Shimizu H., Li H., Sundaramoorthy M., McRee D.E., Goodin D.B., and Poulos T.L., *High-resolution crystal structures and spectroscopy of native and compound I cytochrome c peroxidase.* Biochemistry, 2003. **42**(19): p. 5600-5608.
149. Casadei C.M., Gumiero A., Metcalfe C.L., Murphy E.J., Basran J., Concilio M.G., Teixeira S.C., Schrader T.E., Fielding A.J., Ostermann A., Blakeley M.P., Raven E.L., and Moody P.C., *Heme enzymes. Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase.* Science, 2014. **345**(6193): p. 193-197.
150. Schulz C.E., Rutter R., Sage J.T., Debrunner P.G., and Hager L.P., *Mossbauer and electron paramagnetic resonance studies of horseradish peroxidase and its catalytic intermediates.* Biochemistry, 1984. **23**(20): p. 4743-4754.
151. Yonetani T. and Schleyer H., *Studies on cytochrome c peroxidase. IX. The reaction of ferrimyoglobin with hydroperoxides and a comparison of peroxide-induced compounds of ferrimyoglobin and cytochrome c peroxidase.* J Biol Chem, 1967. **242**(8): p. 1974-1979.
152. Yonetani T. and Schleyer H., *Studies on cytochrome c peroxidase. VII. Electron paramagnetic resonance absorptions of the enzyme and complex ES in dissolved and crystalline forms.* J Biol Chem, 1966. **241**(13): p. 3240-3243.
153. Sivaraja M., Goodin D.B., Smith M., and Hoffman B.M., *Identification by ENDOR of Trp191 as the free-radical site in cytochrome c peroxidase compound ES.* Science, 1989. **245**(4919): p. 738-740.
154. Erman J.E., Vitello L.B., Mauro J.M., and Kraut J., *Detection of an oxyferryl porphyrin π -*

- cation-radical intermediate in the reaction between hydrogen peroxide and a mutant yeast cytochrome c peroxidase. Evidence for tryptophan-191 involvement in the radical site of compound I.* Biochemistry, 1989. **28**(20): p. 7992-7995.
155. Miller M.A., Han G.W., and Kraut J., *A cation binding motif stabilizes the compound I radical of cytochrome c peroxidase.* Proc Natl Acad Sci USA, 1994. **91**(23): p. 11118-11122.
 156. Mauro J.M., Fishel L.A., Hazzard J.T., Meyer T.E., Tollin G., Cusanovich M.A., and Kraut J., *Tryptophan-191 → phenylalanine, a proximal-side mutation in yeast cytochrome c peroxidase that strongly affects the kinetics of ferrocyanochrome c oxidation.* Biochemistry, 1988. **27**(17): p. 6243-6256.
 157. Huyett J.E., Doan P.E., Gurbiel R., Houseman A.L.P., Sivaraja M., Goodin D.B., and Hoffman B.M., *Compound ES of cytochrome c peroxidase contains a Trp π -cation radical - Characterization by continuous wave and pulsed Q-band external nuclear double resonance spectroscopy.* J Am Chem Soc, 1995. **117**(35): p. 9033-9041.
 158. Fitzgerald M.M., Churchill M.J., Mceee D.E., and Goodin D.B., *Small-molecule binding to an artificially created cavity at the active-site of cytochrome c peroxidase.* Biochemistry, 1994. **33**(13): p. 3807-3818.
 159. Bernini C., Arezzini E., Basosi R., and Sinicropi A., *In silico spectroscopy of tryptophan and tyrosine radicals involved in the long-range electron transfer of cytochrome c peroxidase.* J Phys Chem B, 2014. **118**(32): p. 9525-9537.
 160. Hahm S., Geren L., Durham B., and Millett F., *Reaction of cytochrome c with the radical in cytochrome c peroxidase compound I.* J Am Chem Soc, 1993. **115**(8): p. 3372-3373.
 161. Hahm S., Miller M.A., Geren L., Kraut J., Durham B., and Millett F., *Reaction of horse cytochrome c with the radical and the oxyferryl heme in cytochrome c peroxidase compound I.* Biochemistry, 1994. **33**(6): p. 1473-1480.
 162. Miller M.A., Liu R.Q., Hahm S., Geren L., Hibdon S., Kraut J., Durham B., and Millett F., *Interaction domain for the reaction of cytochrome c with the radical and the oxyferryl heme in cytochrome c peroxidase compound I.* Biochemistry, 1994. **33**(29): p. 8686-8693.
 163. Wang K.F., Mei H.K., Geren L., Miller M.A., Saunders A., Wang X.M., Waldner J.L., Pielak G.J., Durham B., and Millett F., *Design of a ruthenium-cytochrome c derivative to measure electron transfer to the radical cation and oxyferryl heme in cytochrome c peroxidase.* Biochemistry, 1996. **35**(47): p. 15107-15119.
 164. Liu R.Q., Miller M.A., Han G.W., Hahm S., Geren L., Hibdon S., Kraut J., Durham B., and Millett F., *Role of methionine-230 in intramolecular electron-transfer between the oxyferryl heme and tryptophan-191 in cytochrome c peroxidase compound II.* Biochemistry, 1994. **33**(29): p. 8678-8685.
 165. Liu R.Q., Geren L., Anderson P., Fairris J.L., Peffer N., McKee A., Durham B., and Millett F., *Design of ruthenium-cytochrome c derivatives to measure electron transfer to cytochrome c peroxidase.* Biochimie, 1995. **77**(7-8): p. 549-561.
 166. Geren L., Hahm S., Durham B., and Millett F., *Photoinduced electron transfer between cytochrome c peroxidase and yeast cytochrome c labeled at Cys 102 with (4-bromomethyl-4'-methylbipyridine)[bis(bipyridine)]ruthenium²⁺.* Biochemistry, 1991. **30**(39): p. 9450-9457.
 167. Hahm S., Durham B., and Millett F., *Photoinduced electron transfer between cytochrome c peroxidase and horse cytochrome c labeled at specific lysines with (dicarboxybipyridine)(bisbipyridine)ruthenium(II).* Biochemistry, 1992. **31**(13): p. 3472-3477.
 168. Mei H., Wang K., McKee S., Wang X., Waldner J.L., Pielak G.J., Durham B., and Millett F., *Control of formation and dissociation of the high-affinity complex between cytochrome c and cytochrome c peroxidase by ionic strength and the low-affinity binding site.* Biochemistry, 1996. **35**(49): p. 15800-15806.

References

169. Pelletier H. and Kraut J., *Crystal-structure of a complex between electron-transfer partners, cytochrome c peroxidase and cytochrome c*. Science, 1992. **258**(5089): p. 1748-1755.
170. Bashir Q., Meulenbroek E.M., Pannu N.S., and Ubbink M., *Engineering specificity in a dynamic protein complex with a single conserved mutation*. FEBS J, 2014. **281**(21): p. 4892-4905.
171. Erman J.E., Kresheck G.C., Vitello L.B., and Miller M.A., *Cytochrome c cytochrome c peroxidase complex: Effect of binding-site mutations on the thermodynamics of complex formation*. Biochemistry, 1997. **36**(13): p. 4054-4060.
172. Pielak G.J. and Wang X.M., *Interactions between yeast iso-1-cytochrome c and its peroxidase*. Biochemistry, 2001. **40**(2): p. 422-428.
173. Hake R., Mclendon G., Corin A., and Holzschu D., *Redox-dependent molecular recognition in proteins - Site-directed mutagenesis suggests that cytochrome c oxidation state governs binding and recognition to cytochrome c peroxidase*. J Am Chem Soc, 1992. **114**(13): p. 5442-5443.
174. Corin A.F., Mclendon G., Zhang Q.P., Hake R.A., Falvo J., Lu K.S., Ciccarelli R.B., and Holzschu D., *Effects of surface amino-acid replacements in cytochrome c peroxidase on complex formation with cytochrome c*. Biochemistry, 1991. **30**(49): p. 11585-11595.
175. Corin A.F., Hake R.A., Mclendon G., Hazzard J.T., and Tollin G., *Effects of surface amino-acid replacements in cytochrome c peroxidase on intracomplex electron-transfer from cytochrome c*. Biochemistry, 1993. **32**(11): p. 2756-2762.
176. Miller M.A., *A complete mechanism for steady-state oxidation of yeast cytochrome c by yeast cytochrome c peroxidase*. Biochemistry, 1996. **35**(49): p. 15791-15799.
177. Liu R.Q., Hahn S., Miller M., Durham B., and Millett F., *Photooxidation of Trp-191 in cytochrome c peroxidase by ruthenium cytochrome c derivatives*. Biochemistry, 1995. **34**(3): p. 973-983.
178. Pearl N.M., Jacobson T., Arisa M., Vitello L.B., and Erman J.E., *Effect of single-site charge-reversal mutations on the catalytic properties of yeast cytochrome c peroxidase: Mutations near the high-affinity cytochrome c binding site*. Biochemistry, 2007. **46**(28): p. 8263-8272.
179. Pearl N.M., Jacobson T., Meyen C., Clementz A.G., Ok E.Y., Choi E., Wilson K., Vitello L.B., and Erman J.E., *Effect of single-site charge-reversal mutations on the catalytic properties of yeast cytochrome c peroxidase: Evidence for a single, catalytically active, cytochrome c binding domain*. Biochemistry, 2008. **47**(9): p. 2766-2775.
180. Erman J.E., Vitello L.B., Pearl N.M., Jacobson T., Francis M., Alberts E., Kou A., and Bujarska K., *Binding of yeast cytochrome c to forty-four charge-reversal mutants of yeast cytochrome c peroxidase: isothermal titration calorimetry*. Biochemistry, 2015.
181. Miller M.A., Geren L., Han G.W., Saunders A., Beasley J., Pielak G.J., Durham B., Millett F., and Kraut J., *Identifying the physiological electron transfer site of cytochrome c peroxidase by structure-based engineering*. Biochemistry, 1996. **35**(18): p. 5948-5948.
182. Kang S.A. and Crane B.R., *Effects of interface mutations on association modes and electron-transfer rates between proteins*. Proc Natl Acad Sci USA, 2005. **102**(43): p. 15465-15470.
183. Worrall J.A., Kolczak U., Canters G.W., and Ubbink M., *Interaction of yeast iso-1-cytochrome c with cytochrome c peroxidase investigated by [¹⁵N, ¹H] heteronuclear NMR spectroscopy*. Biochemistry, 2001. **40**(24): p. 7069-7076.
184. Moench S.J., Chroni S., Lou B.S., Erman J.E., and Satterlee J.D., *Proton NMR comparison of noncovalent and covalently cross-linked complexes of cytochrome c peroxidase with horse, tuna, and yeast ferricytochromes c*. Biochemistry, 1992. **31**(14): p. 3661-3670.

185. Liang N., Mauk A.G., Pielak G.J., Johnson J.A., Smith M., and Hoffman B.M., Regulation of interprotein electron transfer by residue 82 of yeast cytochrome *c*. *Science*, 1988. **240**(4850): p. 311-313.
186. Liang N., Pielak G.J., Mauk A.G., Smith M., and Hoffman B.M., Yeast cytochrome *c* with phenylalanine or tyrosine at position 87 transfers electrons to (zinc cytochrome *c* peroxidase)⁺ at a rate ten thousand times that of the serine-87 or glycine-87 variants. *Proc Natl Acad Sci USA*, 1987. **84**(5): p. 1249-1252.
187. Das G., Hickey D.R., Principio L., Conklin K.T., Short J., Miller J.R., McLendon G., and Sherman F., Replacements of lysine 32 in yeast cytochrome *c*. Effects on the binding and reactivity with physiological partners. *J Biol Chem*, 1988. **263**(34): p. 18290-18297.
188. Kang C.H., Fergusonmiller S., and Margoliash E., Steady-state kinetics and binding of eukaryotic cytochromes *c* with yeast cytochrome *c* peroxidase. *J Biol Chem*, 1977. **252**(3): p. 919-926.
189. Kang D.S. and Erman J.E., The cytochrome *c* peroxidase-catalyzed oxidation of ferrocyanide by hydrogen peroxide. Steady state kinetic mechanism. *J Biol Chem*, 1982. **257**(21): p. 12775-12779.
190. Kornblatt J.A. and English A.M., The binding of porphyrin cytochrome *c* to yeast cytochrome *c* peroxidase - A fluorescence study of the number of sites and their sensitivity to salt. *Eur J Biochem*, 1986. **155**(3): p. 505-511.
191. Matthis A.L., Vitello L.B., and Erman J.E., Oxidation of yeast iso-1-ferrocyanide by yeast cytochrome *c* peroxidase compound I and compound II dependence upon ionic strength. *Biochemistry*, 1995. **34**(31): p. 9991-9999.
192. Stemp E.D.A. and Hoffman B.M., Cytochrome *c* peroxidase binds two molecules of cytochrome *c* - Evidence for a low-affinity, electron-transfer-active site on cytochrome *c* peroxidase. *Biochemistry*, 1993. **32**(40): p. 10848-10865.
193. Zhou J.S. and Hoffman B.M., Cytochrome *c* peroxidase simultaneously binds cytochrome *c* at two different sites with strikingly different reactivities - Titrating a substrate with an enzyme. *J Am Chem Soc*, 1993. **115**(23): p. 11008-11009.
194. Zhou J.S. and Hoffman B.M., Stern-Volmer in reverse - 2:1 stoichiometry of the cytochrome *c* cytochrome *c* peroxidase electron-transfer complex. *Science*, 1994. **265**(5179): p. 1693-1696.
195. Mauk M.R., Ferrer J.C., and Mauk A.G., Proton linkage in formation of the cytochrome *c* - cytochrome *c* peroxidase complex - Electrostatic properties of the high-affinity and low-affinity cytochrome binding sites on the peroxidase. *Biochemistry*, 1994. **33**(42): p. 12609-12614.
196. Zhou J.S., Nocek J.M., DeVan M.L., and Hoffman B.M., Inhibitor-enhanced electron transfer: copper cytochrome *c* as a redox-inert probe of ternary complexes. *Science*, 1995. **269**(5221): p. 204-207.
197. Leesch V.W., Bujons J., Mauk A.G., and Hoffman B.M., Cytochrome *c* peroxidase-cytochrome *c* complex: locating the second binding domain on cytochrome *c* peroxidase with site-directed mutagenesis. *Biochemistry*, 2000. **39**(33): p. 10132-10139.
198. Nocek J.M., Zhou J.S., De Forest S., Priyadarshy S., Beratan D.N., Onuchic J.N., and Hoffman B.M., Theory and practice of electron transfer within protein-protein complexes: Application to the multidomain binding of cytochrome *c* by cytochrome *c* peroxidase. *Chem Rev*, 1996. **96**(7): p. 2459-2490.
199. Morar A.S. and Pielak G.J., Crowding by trisaccharides and the 2:1 cytochrome *c*-cytochrome *c* peroxidase complex. *Biochemistry*, 2002. **41**(2): p. 547-551.
200. Van de Water K., Sterckx Y.G.J., and Volkov A.N., The low-affinity complex of cytochrome *c* and its peroxidase. *Nat Commun*, 2015. **6**(7073): p. 1-12.
201. Page T.R. and Hoffman B.M., Control of cyclic photoinitiated electron transfer between cytochrome *c*

References

- peroxidase (W191F) and cytochrome c by formation of dynamic binary and ternary complexes.* *Biochemistry*, 2015. **54**(5): p. 1188-1197.
202. Jeng M.F., Englander S.W., Pardue K., Rogalsky J.S., and McLendon G., *Structural dynamics in an electron-transfer complex.* *Nat Struc Biol*, 1994. **1**(4): p. 234-238.
203. Yi Q., Erman J.E., and Satterlee J.D., *Studies of protein-protein association between yeast cytochrome c peroxidase and yeast iso-1 ferricytochrome c by hydrogen-deuterium exchange labeling and proton NMR spectroscopy.* *Biochemistry*, 1994. **33**(40): p. 12032-12041.
204. Van de Water K., van Nuland N.A.J., and Volkov A.N., *Transient protein encounters characterized by paramagnetic NMR.* *Chem Sci*, 2014. **5**(11): p. 4227-4236.
205. McLendon G., Wallin S.A., Miller R.M., Billstone V., Spears K.G., Hoffman B.M., and Zhang Q., *Thermodynamic and kinetic aspects of binding and recognition in the cytochrome c/cytochrome c peroxidase complex.* *J Am Chem Soc*, 1993. **115**(9): p. 3665-3669.
206. Yi Q., Erman J.E., and Satterlee J.D., *¹H NMR evaluation of yeast isozyme-1 ferricytochrome c equilibrium exchange dynamics in noncovalent complexes with two forms of yeast cytochrome c peroxidase.* *J Am Chem Soc*, 1994. **116**(5): p. 1981-1987.
207. Mei H., Geren L., Miller M.A., Durham B., and Millett F., *Role of the low-affinity binding site in electron transfer from cytochrome c to cytochrome c peroxidase.* *Biochemistry*, 2002. **41**(12): p. 3968-3976.
208. Pappa H.S. and Poulos T.L., *Site-specific cross-linking as a method for studying intramolecular electron transfer.* *Biochemistry*, 1995. **34**(20): p. 6573-6580.
209. Pappa H.S., Tajbaksh S., Saunders A.J., Pielak G.J., and Poulos T.L., *Probing the cytochrome c peroxidase-cytochrome c electron transfer reaction using site specific cross-linking.* *Biochemistry*, 1996. **35**(15): p. 4837-4845.
210. Guo M., Bhaskar B., Li H., Barrows T.P., and Poulos T.L., *Crystal structure and characterization of a cytochrome c peroxidase-cytochrome c site-specific cross-link.* *Proc Natl Acad Sci USA*, 2004. **101**(s16): p. 5940-5945.
211. Nakani S., Viriyakul T., Mitchell R., Vitello L.B., and Erman J.E., *Characterization of a covalently linked yeast cytochrome c-cytochrome c peroxidase complex: evidence for a single, catalytically active cytochrome c binding site on cytochrome c peroxidase.* *Biochemistry*, 2006. **45**(32): p. 9887-9893.
212. Seifert J.L., Pfister T.D., Nocek J.M., Lu Y., and Hoffman B.M., *Hopping in the electron-transfer photocycle of the 1:1 complex of Zn-cytochrome c peroxidase with cytochrome c.* *J Am Chem Soc*, 2005. **127**(16): p. 5750-5751.
213. Kang S.A., Marjavaara P.J., and Crane B.R., *Electron transfer between cytochrome c and cytochrome c peroxidase in single crystals.* *J Am Chem Soc*, 2004. **126**(35): p. 10836-10837.
214. Erman J.E. and Vitello L.B., *Yeast cytochrome c peroxidase: mechanistic studies via protein engineering.* *Bba-Protein Struct M*, 2002. **1597**(2): p. 193-220.
215. Yonetani T., *Studies on cytochrome c peroxidase. X. Crystalline apo- and reconstituted holoenzymes.* *J Biol Chem*, 1967. **242**(21): p. 5008-5013.
216. Vanwetswinkel S., van Nuland N.A.J., and Volkov A.N., *Paramagnetic properties of the low- and high-spin states of yeast cytochrome c peroxidase.* *J Biomol NMR*, 2013. **57**(1): p. 21-26.
217. Volkov A.N. and van Nuland N.A., *Solution NMR study of the yeast cytochrome c peroxidase: cytochrome c interaction.* *J Biomol NMR*, 2008. **56**(3): p. 255-263.
218. Van de Water K., Sterckx Y.G., and Volkov A.N., *The low-affinity complex of cytochrome c and its peroxidase.* *Nat Commun*, 2015. **6**: p. 7073.
219. Van de Water K., van Nuland N.A.J., and Volkov A.N., *Transient protein encounters*

- characterized by paramagnetic NMR. *Chemical Science*, 2014. **5**(11): p. 4227-4236.
220. Vitello L.B., Huang M., and Erman J.E., *pH-dependent spectral and kinetic properties of cytochrome c peroxidase: comparison of freshly isolated and stored enzyme*. *Biochemistry*, 1990. **29**(18): p. 4283-4288.
 221. Nelson C.E., Sitzman E.V., Kang C.H., and Margoliash E., *Preparation of cytochrome c peroxidase from baker's yeast*. *Anal Biochem*, 1977. **83**(2): p. 622-631.
 222. English A.M., Laberge M., and Walsh M., *Rapid procedure for the isolation of cytochrome c peroxidase*. *Inorg Chim Acta*, 1986. **123**(2): p. 113-116.
 223. Smulevich G., Evangelista-Kirkup R., English A., and Spiro T.G., *Raman and infrared spectra of cytochrome c peroxidase-carbon monoxide adducts in alternative conformational states*. *Biochemistry*, 1986. **25**(15): p. 4426-4430.
 224. Dasgupta S., Rousseau D.L., Anni H., and Yonetani T., *Structural characterization of cytochrome c peroxidase by resonance Raman scattering*. *J Biol Chem*, 1989. **264**(1): p. 654-662.
 225. Satterlee J.D. and Erman J.E., *Proton magnetic-resonance studies of cytochrome c peroxidase - pD dependence of the isotropically shifted resonances*. *Arch Biochem Biophys*, 1980. **202**(2): p. 608-616.
 226. Wang J.L., Boldt N.J., and Ondrias M.R., *Formation and photolability of low-spin ferrous cytochrome c peroxidase at alkaline pH*. *Biochemistry*, 1992. **31**(3): p. 867-878.
 227. Conroy C.W., Tyma P., Daum P.H., and Erman J.E., *Oxidation-reduction potential measurements of cytochrome c peroxidase and pH dependent spectral transitions in the ferrous enzyme*. *Biochim Biophys Acta*, 1978. **537**(1): p. 62-69.
 228. Shelnutt J.A., Satterlee J.D., and Erman J.E., *Raman difference spectroscopy of heme-linked ionizations in cytochrome c peroxidase*. *J Biol Chem*, 1983. **258**(4): p. 2168-2173.
 229. Hashimoto S., Teraoka J., Inubushi T., Yonetani T., and Kitagawa T., *Resonance Raman study on cytochrome c peroxidase and its intermediate. Presence of the Fe(IV) = O bond in compound E.S and heme-linked ionization*. *J Biol Chem*, 1986. **261**(24): p. 11110-11118.
 230. Loo S. and Erman J.E., *Kinetic study of reaction between cytochrome c peroxidase and hydrogen peroxide dependence on pH and ionic strength*. *Biochemistry*, 1975. **14**(15): p. 3467-3470.
 231. Lent B., Conroy C.W., and Erman J.E., *Effect of ionic strength on kinetics of fluoride binding to cytochrome c peroxidase*. *Arch Biochem Biophys*, 1976. **177**(1): p. 56-61.
 232. Coulson A.F., Erman J.E., and Yonetani T., *Studies on cytochrome c peroxidase. XVII. Stoichiometry and mechanism of the reaction of compound E.S with donors*. *J Biol Chem*, 1971. **246**(4): p. 917-924.
 233. Balny C., Saldana J.L., and Dahan N., *High-pressure stopped-flow spectrometry at low temperatures*. *Anal Biochem*, 1984. **139**(1): p. 178-189.
 234. Miller M.A., Mauro J.M., Smulevich G., Coletta M., Kraut J., and Traylor T.G., *CO dissociation in cytochrome c peroxidase: site-directed mutagenesis shows that distal Arg 48 influences CO dissociation rates*. *Biochemistry*, 1990. **29**(42): p. 9978-9988.
 235. Miller M.A., Coletta M., Mauro J.M., Putnam L.D., Farnum M.F., Kraut J., and Traylor T.G., *CO recombination in cytochrome c peroxidase: effect of the local heme environment on CO binding explored through site-directed mutagenesis*. *Biochemistry*, 1990. **29**(7): p. 1777-1791.
 236. Smulevich G., Miller M.A., Kraut J., and Spiro T.G., *Conformational change and histidine control of heme chemistry in cytochrome c peroxidase*:

References

- resonance Raman evidence from *Leu-52* and *Gly-181* mutants of cytochrome *c* peroxidase. *Biochemistry*, 1991. **30**(39): p. 9546-9558.
237. Englander S.W. and Kallenbach N.R., *Hydrogen exchange and structural dynamics of proteins and nucleic acids*. *Q Rev Biophys*, 1983. **16**(4): p. 521-655.
238. Weigelt J., *Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments*. *J Am Chem Soc*, 1998. **120**(41): p. 10778-10779.
239. Czisch M. and Boelens R., *Sensitivity enhancement in the TROSY experiment*. *J Magn Reson*, 1998. **134**(1): p. 158-160.
240. Pervushin K.V., Wider G., and Wuthrich K., *Single transition-to-single transition polarization transfer (ST2-PT) in [¹⁵N,¹H]-TROSY*. *J Biomol NMR*, 1998. **12**(2): p. 345-348.
241. Meissner A., Schulte-Herbruggen T., Briand J., and Sorensen O.W., *Double spin-state-selective coherence transfer. Application for two-dimensional selection of multiplet components with long transverse relaxation times*. *Mol Phys*, 1998. **95**(6): p. 1137-1142.
242. Vranken W.F., Boucher W., Stevens T.J., Fogh R.H., Pajon A., Llinas M., Ulrich E.L., Markley J.L., Ionides J., and Laue E.D., *The CCPN data model for NMR spectroscopy: development of a software pipeline*. *Proteins*, 2005. **59**(4): p. 687-696.
243. Salzmann M., Wider G., Pervushin K., Senn H., and Wuthrich K., *TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins*. *J Am Chem Soc*, 1999. **121**(4): p. 844-848.
244. Favier A. and Brutscher B., *Recovering lost magnetization: polarization enhancement in biomolecular NMR*. *J Biomol NMR*, 2011. **49**(1): p. 9-15.
245. Lescop E., Kern T., and Brutscher B., *Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: Example of longitudinal-relaxation-enhanced BEST-type ¹H-¹⁵N correlation experiments*. *J Magn Reson*, 2010. **203**(1): p. 190-198.
246. Tsan P., Caffrey M., Daku M.L., Cusanovich M., Marion D., and Gans P., *Magnetic susceptibility tensor and heme contact shifts determinations in the *Rhodobacter capsulatus* ferricytochrome *c*: NMR and magnetic susceptibility studies*. *J Am Chem Soc*, 2001. **123**(10): p. 2231-2242.
247. Inubushi T. and Becker E.D., *Efficient detection of paramagnetically shifted NMR resonances by optimizing the WFT pulse sequence*. *J Magn Reson*, 1983. **51**(1): p. 128-133.
248. Studier F.W., Rosenberg A.H., Dunn J.J., and Dubendorff J.W., *Use of T7 RNA polymerase to direct expression of cloned genes*. *Methods Enzymol*, 1990. **185**: p. 60-89.
249. Dubendorff J.W. and Studier F.W., *Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor*. *J Mol Biol*, 1991. **219**(1): p. 45-59.
250. Studier F.W. and Moffatt B.A., *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes*. *J Mol Biol*, 1986. **189**(1): p. 113-130.
251. Davanloo P., Rosenberg A.H., Dunn J.J., and Studier F.W., *Cloning and expression of the gene for bacteriophage T7 RNA polymerase*. *Proc Natl Acad Sci USA*, 1984. **81**(7): p. 2035-2039.
252. Moffatt B.A. and Studier F.W., *T7 lysozyme inhibits transcription by T7 RNA polymerase*. *Cell*, 1987. **49**(2): p. 221-227.
253. Kido M., Yamanaka K., Mitani T., Niki H., Ogura T., and Hiraga S., *RNase E polypeptides lacking a carboxyl-terminal half suppress a *mukB* mutation in *Escherichia coli**. *J Bacteriol*, 1996. **178**(13): p. 3917-3925.

254. Lopez P.J., Marchand I., Joyce S.A., and Dreyfus M., *The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo.* Mol Microbiol, 1999. **33**(1): p. 188-199.
255. Schilder J., Löhr F., Schwalbe H., and Ubbink M., *Corrigendum to "The cytochrome c peroxidase and cytochrome c encounter complex: The other side of the story" [FEBS Lett. 588 (2014) 1873–1878].* FEBS Letters, 2015. **589**(16): p. 2169.
256. Ottiger M., Tjandra N., and Bax A., *Magnetic field dependent amide ¹⁵N chemical shifts in a protein-DNA complex resulting from magnetic ordering in solution.* J Am Chem Soc, 1997. **119**(41): p. 9825-9830.
257. John M., Park A.Y., Pintacuda G., Dixon N.E., and Otting G., *Weak alignment of paramagnetic proteins warrants correction for residual CSA effects in measurements of pseudocontact shifts.* J Am Chem Soc, 2005. **127**(49): p. 17190-17191.
258. Vallurupalli P., Hansen D.F., Stollar E., Meirovitch E., and Kay L.E., *Measurement of bond vector orientations in invisible excited states of proteins.* Proc Natl Acad Sci USA, 2007. **104**(47): p. 18473-18477.
259. Budd D.L., Lamar G.N., Langry K.C., Smith K.M., and Nayyirmazhir R., *¹H NMR study of high-spin ferric natural porphyrin derivatives as models of metbemoproteins.* J Am Chem Soc, 1979. **101**(20): p. 6091-6096.
260. Sterckx Y.G. and Volkov A.N., *Cofactor-dependent structural and binding properties of yeast cytochrome c peroxidase.* Biochemistry, 2014. **53**(28): p. 4526-4236.
261. Ubbink M., Lian L.Y., Modi S., Evans P.A., and Bendall D.S., *Analysis of the ¹H-NMR chemical shifts of Cu(I)-, Cu(II)- and Cd-substituted pea plastocyanin. Metal-dependent differences in the hydrogen-bond network around the copper site.* Eur J Biochem, 1996. **242**(1): p. 132-147.
262. Yonetani T. and Ray G.S., *Studies on cytochrome c peroxidase. I. Purification and some properties.* J Biol Chem, 1965. **240**(11): p. 4503-4508.
263. Yonetani T., Chance B., and Kajiwara S., *Crystalline cytochrome c peroxidase and complex ES.* J Biol Chem, 1966. **241**(12): p. 2981-2982.
264. Walker F.A. and La Mar G.N., *Nuclear magnetic resonance studies of ferric porphyrins.* Ann N Y Acad Sci, 1973. **206**: p. 328-348.
265. Andersson K.K., Lipscomb J.D., Valentine M., Munck E., and Hooper A.B., *Tetraheme cytochrome c-554 from Nitrosomonas europaea. Heme-heme interactions and ligand binding.* J Biol Chem, 1986. **261**(3): p. 1126-1138.
266. Kao Y.H. and Lecomte J.T.J., *Determination of the zero-field splitting constant for proton NMR chemical-shift analysis in metaquomoglobin - the dipolar shift as a structural probe.* J Am Chem Soc, 1993. **115**(21): p. 9754-9762.
267. Brackett G.C., Richards P.L., and Caughey W.S., *Far-infrared magnetic resonance in Fe(III) and Mn(III) porphyrins, myoglobin, hemoglobin, ferrichrome a, and Fe(III) dithiocarbamates.* J Chem Phys, 1971. **54**(10): p. 4383-4401.
268. Scholes C.P., Isaacson R.A., and Feher G., *Determination of zero-field splitting of Fe³⁺ in heme proteins from temperature dependence of spin-lattice relaxation rate.* Biochim Biophys Acta, 1971. **244**(1): p. 206-210.
269. Rajarathnam K., Lamar G.N., Chiu M.L., Sligar S.G., Singh J.P., and Smith K.M., *¹H NMR hyperfine shift pattern as a probe for ligation state in high-spin ferric hemoproteins - water binding in metmyoglobin mutants.* J Am Chem Soc, 1991. **113**(21): p. 7886-7892.
270. Bou-Abdallah F. and Chasteen N.D., *Spin concentration measurements of high-spin (g' = 4.3) rhombic iron(III) ions in biological samples: theory and application.* J Biol Inorg Chem, 2008. **13**(1): p. 15-24.

References

271. Peisach J., Blumberg W.E., Ogawa S., Rachmilewitz E.A., and Oltzik R., *The effects of protein conformation on the heme symmetry in high spin ferric heme proteins as studied by electron paramagnetic resonance.* J Biol Chem, 1971. **246**(10): p. 3342-3355.
272. McDonald G.R., Hudson A.L., Dunn S.M., You H., Baker G.B., Whittal R.M., Martin J.W., Jha A., Edmondson D.E., and Holt A., *Bioactive contaminants leach from disposable laboratory plasticware.* Science, 2008. **322**(5903): p. 917.
273. Gardner R.J., Longinetti M., and Sgheri L., *Reconstruction of orientations of a moving protein domain from paramagnetic data.* Inverse Probl, 2005. **21**(3): p. 879-898.
274. Longinetti M., Luchinat C., Parigi G., and Sgheri L., *Efficient determination of the most favoured orientations of protein domains from paramagnetic NMR data.* Inverse Probl, 2006. **22**(4): p. 1485-1502.
275. Bertini I., Luchinat C., Nagulapalli M., Parigi G., and Ravera E., *Paramagnetic relaxation enhancement for the characterization of the conformational heterogeneity in two-domain proteins.* Phys Chem Chem Phys, 2012. **14**(25): p. 9149-9156.
276. Luchinat C., Nagulapalli M., Parigi G., and Sgheri L., *Maximum occurrence analysis of protein conformations for different distributions of paramagnetic metal ions within flexible two-domain proteins.* J Magn Reson, 2012. **215**: p. 85-93.
277. Bertini I., Gupta Y.K., Luchinat C., Parigi G., Peana M., Sgheri L., and Yuan J., *Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains.* J Am Chem Soc, 2007. **129**(42): p. 12786-12794.
278. Gabdoulline R.R. and Wade R.C., *Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations.* J Mol Biol, 2001. **306**(5): p. 1139-1155.
279. Lescop E., Schanda P., and Brutscher B., *A set of BEST triple-resonance experiments for time-optimized protein resonance assignment.* J Magn Reson, 2007. **187**(1): p. 163-169.
280. Grzesiek S., Bax A., Clore G.M., Gronenborn A.M., Hu J.S., Kaufman J., Palmer I., Stahl S.J., and Wingfield P.T., *The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase.* Nat Struct Biol, 1996. **3**(4): p. 340-345.
281. Kannt A., Young S., and Bendall D.S., *The role of acidic residues of plastocyanin in its interaction with cytochrome *f*.* BBA-Bioenergetics, 1996. **1277**(1-2): p. 115-126.
282. Schwieters C.D., Kuszewski J.J., and Clore G.M., *Using Xplor-NIH for NMR molecular structure determination.* Prog Nucl Mag Res Sp, 2006. **48**(1): p. 47-62.
283. Schwieters C.D., Kuszewski J.J., Tjandra N., and Clore G.M., *The Xplor-NIH NMR molecular structure determination package.* J Magn Reson, 2003. **160**(1): p. 65-73.
284. Hervas M., Bashir Q., Leferink N.G.H., Ferreira P., Moreno-Beltran B., Westphal A.H., Diaz-Moreno I., Medina M., de la Rosa M.A., Ubbink M., Navarro J.A., and van Berkel W.J.H., *Communication between L-galactono-1,4-lactone dehydrogenase and cytochrome *c*.* FEBS J, 2013. **280**(8): p. 1830-1840.
285. Griffith O.H. and McConnell Hm, *A nitroxide-maleimide spin label.* Proc Natl Acad Sci U S A, 1966. **55**(1): p. 8-11.
286. Jeschke G., *Conformational dynamics and distribution of nitroxide spin labels.* Prog Nucl Mag Res Sp, 2013. **72**: p. 42-60.
287. Bermejo G.A., Strub M.P., Ho C., and Tjandra N., *Determination of the solution-bound conformation of an amino acid binding protein by NMR paramagnetic relaxation enhancement: use of a single flexible paramagnetic probe with improved*

- estimation of its sampling space. *J Am Chem Soc*, 2009. **131**(27): p. 9532-9537.
288. Tang C., Schwieters C.D., and Clore G.M., *Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR*. *Nature*, 2007. **449**(7165): p. 1078-1082.
289. Iwahara J., Schwieters C.D., and Clore G.M., *Characterization of nonspecific protein-DNA interactions by ¹H paramagnetic relaxation enhancement*. *J Am Chem Soc*, 2004. **126**(40): p. 12800-12808.
290. Clore G.M., Tang C., and Iwahara J., *Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement*. *Curr Opin Struct Biol*, 2007. **17**(5): p. 603-616.
291. Fawzi N.L., Fleissner M.R., Anthis N.J., Kalai T., Hideg K., Hubbell W.L., and Clore G.M., *A rigid disulfide-linked nitroxide side chain simplifies the quantitative analysis of PRE data*. *J Biomol NMR*, 2011. **51**(1-2): p. 105-114.
292. Stoll S. and Schweiger A., *EasySpin, a comprehensive software package for spectral simulation and analysis in EPR*. *J Magn Reson*, 2006. **178**(1): p. 42-55.
293. Steigmiller S., Börsch M., Gräber P., and Huber M., *Distances between the b-subunits in the tether domain of F₀F₁-ATP synthase from E. coli*. *Biochim Biophys Acta - Bioenerg*, 2005. **1708**(2): p. 143-153.
294. Gao Y., Boyd J., Williams R.J., and Pielak G.J., *Assignment of proton resonances, identification of secondary structural elements, and analysis of backbone chemical shifts for the C102T variant of yeast iso-1-cytochrome c and horse cytochrome c*. *Biochemistry*, 1990. **29**(30): p. 6994-7003.
295. Fetrow J.S. and Baxter S.M., *Assignment of ¹⁵N chemical shifts and ¹⁵N relaxation measurements for oxidized and reduced iso-1-cytochrome c*. *Biochemistry*, 1999. **38**(14): p. 4480-4492.
296. Clore G.M. and Iwahara J., *Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes*. *Chem Rev*, 2009. **109**(9): p. 4108-4139.
297. Bernardo P., Garcia de la Torre J., and Pons M., *Interpretation of ¹⁵N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR*. *J Biomol NMR*, 2002. **23**(2): p. 139-150.
298. Zagrovic B. and Pande V.S., *How does averaging affect protein structure comparison on the ensemble level?* *Biophys J*, 2004. **87**(4): p. 2240-2246.
299. Takayama Y. and Clore G.M., *Intra- and intermolecular translocation of the bi-domain transcription factor Oct1 characterized by liquid crystal and paramagnetic NMR*. *Proc Natl Acad Sci U S A*, 2011. **108**(22): p. E169-E176.
300. Shi L., Traaseth N.J., Verardi R., Gustavsson M., Gao J.L., and Veglia G., *Paramagnetic-based NMR restraints lift residual dipolar coupling degeneracy in multidomain detergent-solubilized membrane proteins*. *J Am Chem Soc*, 2011. **133**(7): p. 2232-2241.
301. Hiruma Y., Hass M.A., Kikui Y., Liu W.M., Olmez B., Skinner S.P., Blok A., Kloosterman A., Koteishi H., Lohr F., Schwalbe H., Nojiri M., and Ubbink M., *The structure of the cytochrome p450cam-putidaredoxin complex determined by paramagnetic NMR spectroscopy and crystallography*. *J Mol Biol*, 2013. **425**(22): p. 4353-4365.
302. Zamoan J., Mascioni A., Thomas D.D., and Veglia G., *NMR solution structure and topological orientation of monomeric phospholamban in dodecylphosphocholine micelles*. *Biophys J*, 2003. **85**(4): p. 2589-2598.
303. Vaynberg J. and Qin J., *Weak protein-protein interactions as probed by NMR spectroscopy*. *Trends Biotechnol* 2006. **24**(1): p. 22-27.

References

304. Tang C., Ghirlando R., and Clore G.M., *Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement*. J Am Chem Soc, 2008. **130**(12): p. 4048-4056.
305. Tang C., Louis J.M., Aniana A., Suh J.Y., and Clore G.M., *Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease*. Nature, 2008. **455**(7213): p. 693-U92.
306. Garcia-Ruiz J.M., *Nucleation of protein crystals*. J Struct Biol, 2003. **142**(1): p. 22-31.
307. Zlotnick A., *Theoretical aspects of virus capsid assembly*. J Mol Recognit, 2005. **18**(6): p. 479-490.
308. Fawzi N.L., Okabe Y., Yap E.H., and Head-Gordon T., *Determining the critical nucleus and mechanism of fibril elongation of the Alzheimer's A beta(1-40) peptide*. J Mol Biol, 2007. **365**(2): p. 535-550.
309. Kleckner I.R. and Foster M.P., *An introduction to NMR-based approaches for measuring protein dynamics*. Biochim Biophys Acta, 2011. **1814**(8): p. 942-968.
310. Suh J.Y., Tang C., and Clore G.M., *Role of electrostatic interactions in transient encounter complexes in protein-protein association investigated by paramagnetic relaxation enhancement*. Journal of the American Chemical Society, 2007. **129**(43): p. 12954-+.
311. Iwahara J., Zweckstetter M., and Clore G.M., *NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA*. Proc Natl Acad Sci U S A, 2006. **103**(41): p. 15062-15067.
312. Henzler-Wildman K.A., Thai V., Lei M., Ott M., Wolf-Watz M., Fenn T., Pozharski E., Wilson M.A., Petsko G.A., Karplus M., Hubner C.G., and Kern D., *Intrinsic motions along an enzymatic reaction trajectory*. Nature, 2007. **450**(7171): p. 838-U13.
313. Shortle D. and Ackerman M.S., *Persistence of native-like topology in a denatured protein in 8 M urea*. Science, 2001. **293**(5529): p. 487-489.
314. Dedmon M.M., Lindorff-Larsen K., Christodoulou J., Vendruscolo M., and Dobson C.M., *Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations*. J Am Chem Soc, 2005. **127**(2): p. 476-477.
315. Gillespie J.R. and Shortle D., *Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures*. J Mol Biol, 1997. **268**(1): p. 170-184.
316. Gillespie J.R. and Shortle D., *Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels*. J Mol Biol, 1997. **268**(1): p. 158-169.
317. Villareal V.A., Spirig T., Robson S.A., Liu M.Y., Lei B.F., and Clubb R.T., *Transient weak protein-protein complexes transfer heme across the cell wall of Staphylococcus aureus*. J Am Chem Soc, 2011. **133**(36): p. 14176-14179.
318. Liu Z., Zhang W.P., Xing Q., Ren X.F., Liu M.L., and Tang C., *Noncovalent dimerization of ubiquitin*. Angew Chem Int Ed Engl, 2012. **51**(2): p. 469-472.
319. Johansson H., Jensen M.R., Gesmar H., Meier S., Vinther J.M., Keeler C., Hodsdon M.E., and Led J.J., *Specific and nonspecific interactions in ultraweak protein-protein associations revealed by solvent paramagnetic relaxation enhancements*. J Am Chem Soc, 2014. **136**(29): p. 10277-10286.
320. Xing Q., Huang P., Yang J., Sun J.Q., Gong Z., Dong X., Guo D.C., Chen S.M., Yang Y.H., Wang Y., Yang M.H., Yi M., Ding Y.M., Liu M.L., Zhang W.P., and Tang C., *Visualizing an ultra-weak protein-protein interaction in phosphorylation signaling*. Angew Chem Int Ed Engl, 2014. **53**(43): p. 11501-11505.

321. Schilder J., Liu W.D., Kumar P., Overhand M., Huber M., and Ubbink M., *Protein docking using an ensemble of spin labels optimized by intra-molecular paramagnetic relaxation enhancement*. Phys Chem Chem Phys, *Accepted*, 2015.
322. Tan K.P., Nguyen T.B., Patel S., Varadarajan R., and Madhusudhan M.S., *DEPTH: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pK(a) of ionizable residues in proteins*. Nucleic Acids Research, 2013. **41**(W1): p. W314-W321.
323. Tan K.P., Varadarajan R., and Madhusudhan M.S., *DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins*. Nucleic Acids Research, 2011. **39**: p. W242-W248.
324. Baker N.A., Sept D., Joseph S., Holst M.J., and McCammon J.A., *Electrostatics of nanosystems: Application to microtubules and the ribosome*. Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-10041.
325. Keizers P.H. and Ubbink M., *Paramagnetic tagging for protein structure and dynamics analysis*. Prog Nucl Magn Reson Spectrosc, 2011. **58**(1-2): p. 88-96.
326. Hass M.A., Keizers P.H., Blok A., Hiruma Y., and Ubbink M., *Validation of a lanthanide tag for the analysis of protein dynamics by paramagnetic NMR spectroscopy*. J Am Chem Soc, 2010. **132**(29): p. 9952-9953.
327. Hass M.A. and Ubbink M., *Structure determination of protein-protein complexes with long-range anisotropic paramagnetic NMR restraints*. Curr Opin Struct Biol, 2014. **24**: p. 45-53.
328. Ulucan O. and Helms V., *How hydrophilic proteins form non-specific complexes*. J Phys Chem B, 2015.
329. Liu W.M., Keizers P.H., Hass M.A., Blok A., Timmer M., Sarris A.J., Overhand M., and Ubbink M., *A pH-sensitive, colorful, lanthanide-chelating paramagnetic NMR probe*. J Am Chem Soc, 2012. **134**(41): p. 17306-13.
330. Piotto M., Saudek V., and Sklenar V., *Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions*. J Biomol NMR, 1992. **2**(6): p. 661-665.
331. Yao L.S., Ying J.F., and Bax A., *Improved accuracy of N¹⁵-H¹ scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins*. J Biomol NMR, 2009. **43**(3): p. 161-170.
332. Volkov A.N., Ferrari D., Worrall J.A., Bonvin A.M., and Ubbink M., *The orientations of cytochrome c in the highly dynamic complex with cytochrome b5 visualized by NMR and docking using HADDOCK*. Protein Sci, 2005. **14**(3): p. 799-811.
333. Bertini I., Luchinat C., and Parigi G., *Magnetic susceptibility in paramagnetic NMR*. Prog Nucl Mag Res Sp, 2002. **40**(3): p. 249-273.
334. Hass M.A., Liu W.M., Agafonov R.V., Otten R., Phung L.A., Schilder J.T., Kern D., and Ubbink M., *A minor conformation of a lanthanide tag on adenylate kinase characterized by paramagnetic relaxation dispersion NMR spectroscopy*. J Biomol NMR, 2015. **61**(2): p. 123-36.
335. Eichmueller C. and Skrynnikov N.R., *Observation of μ s time-scale protein dynamics in the presence of Ln³⁺ ions: application to the N-terminal domain of cardiac troponin C*. J Biomol NMR, 2007. **37**(2): p. 79-95.
336. Liu Z., Gong Z., Dong X., and Tang C., *Transient protein-protein interactions visualized by solution NMR*. Biochim Biophys Acta, 2015.

APPENDIX I

TABLE A1: ^1HN , ^{15}NH , $^{13}\text{C}_\alpha$ and ^{13}CO chemical shifts (in ppm) of the NMR resonance assignments backbone nuclei of the major form observed in the spectra of yeast CcP.

CcP residue	^1HN chemical shift	^{15}NH chemical shift	$^{13}\text{C}_\alpha$ chemical shift	^{13}CO chemical shift
2Lys	8.370	124.054	176.279	58.083
3Thr	8.152	118.545	173.894	55.947
4Leu	8.172	127.297	175.912	61.592
5Val	8.124	124.598	174.519	54.541
6His	8.660	125.105	172.351	60.730
7Val	8.814	125.178	177.808	51.646
8Ala	8.434	135.366	177.153	61.425
9Ser	9.402	121.537	-	52.292
10Val	8.518	128.725	-	56.156
12Lys	8.234	124.547	178.616	58.895
13Gly	8.808	113.431	174.160	58.035
14Arg	7.548	119.000	175.549	45.078
15Ser	9.420	119.866	174.226	52.623
16Tyr	8.931	122.294	176.099	56.468
17Glu	8.832	117.107	179.076	61.474
18Asp	7.491	117.965	177.918	59.457
19Phe	7.401	116.291	177.401	57.401
20Gln	8.820	121.818	177.773	62.304
21Lys	7.359	119.721	178.933	57.330
24Asn	8.741	117.930	176.836	60.424
25Ala	7.816	123.300	-	54.863
27Ala	9.099	122.465	180.688	65.163
28Leu	9.310	121.871	179.609	54.881
29Lys	7.547	123.468	178.572	57.324
30Leu	8.702	120.842	180.068	59.253
31Arg	7.448	117.816	177.822	56.946
32Glu	7.898	118.492	177.926	58.143
33Asp	8.657	123.234	177.558	58.129
34Asp	8.288	119.796	176.496	52.726
35Glu	8.343	118.108	177.220	54.520
36Tyr	6.868	121.603	174.223	56.696
37Asp	8.156	128.374	175.646	58.654
38Asn	8.509	115.230	175.048	53.272
39Tyr	7.490	109.340	175.052	54.556
40Ile	7.975	124.484	179.190	60.176
41Gly	8.763	107.762	173.718	65.876
42Tyr	9.657	116.315	175.507	44.494
43Gly	9.788	109.864	174.904	60.239
53Thr	7.704	117.731	-	57.130
57Trp	6.694	121.881	173.673	59.643
58Asp	7.321	125.958	175.021	58.682
59Lys	6.778	122.955	177.270	52.081
61Asp	6.815	114.494	175.593	58.431
62Asn	8.252	117.690	176.062	51.903
63Thr	7.532	108.636	174.473	53.129
64Gly	8.327	105.636	175.545	60.532
65Gly	7.780	107.799	175.865	44.283
66Ser	8.439	111.976	172.977	43.599
67Tyr	8.324	122.618	177.445	59.525
68Gly	8.294	131.118	175.406	62.474
69Gly	7.581	109.724	175.918	45.912
70Thr	7.397	112.272	177.638	46.856

71Tyr	8.034	124.413	174.040	63.998
72Arg	6.479	108.148	170.824	59.125
73Phe	7.934	123.607	176.257	55.429
74Lys	8.575	122.186	176.799	57.119
75Lys	8.320	117.563	177.245	58.584
77Phe	8.002	124.409	178.736	57.038
78Asn	7.405	112.231	174.047	59.988
79Asp	6.667	124.140	177.233	52.001
82Asn	8.263	127.616	181.281	57.188
84Gly	8.231	114.633	177.549	54.353
86Gln	9.207	122.083	178.231	55.568
87Asn	7.285	116.598	178.321	58.539
88Gly	7.402	107.446	173.517	55.986
89Phe	7.780	123.073	177.473	46.448
90Lys	8.423	117.960	178.954	61.720
91Phe	7.100	121.344	175.718	58.311
92Leu	7.124	114.315	178.441	60.450
93Glu	7.610	123.165	175.909	54.994
95Ile	6.590	119.333	177.555	64.843
96His	8.147	121.159	177.512	62.929
97Lys	7.557	115.519	177.795	55.275
98Glu	7.037	119.551	175.694	58.009
99Phe	7.326	116.942	-	57.139
101Trp	6.611	114.551	177.054	64.438
102Ile	6.832	125.646	173.890	53.687
104Ser	10.072	120.518	174.201	57.228
105Gly	9.344	109.162	178.718	63.327
106Asp	7.520	124.728	176.087	45.673
107Leu	7.763	122.316	178.224	57.731
108Phe	8.532	116.206	-	57.903
110Leu	8.658	127.662	180.830	61.404
111Gly	8.762	109.453	174.790	56.978
112Gly	7.237	105.686	174.616	47.589
115Ala	8.002	121.543	178.660	65.999
119Met	7.827	120.592	173.673	57.979
120Gln	7.655	109.240	176.143	56.391
121Gly	8.277	106.236	-	57.152
123Lys	7.875	112.809	178.804	61.286
124Ile	8.812	125.922	172.337	54.337
126Trp	9.169	125.019	173.011	61.901
127Arg	7.903	126.845	171.937	56.725
128Ala	6.601	121.598	176.852	52.887
131Val	7.001	117.303	172.096	58.934
132Asp	8.047	124.218	177.657	60.524
133Thr	8.505	117.125	173.971	54.490
135Glu	8.559	124.416	178.736	62.651
136Asp	8.357	117.991	176.431	59.596
137Thr	7.838	110.648	174.373	54.956
138Thr	7.341	123.997	172.629	61.620
140Asp	8.057	119.947	175.690	61.726
141Asn	8.367	117.309	174.557	53.916
142Gly	7.659	110.662	175.854	52.920
143Arg	8.523	120.056	175.786	44.581
144Leu	7.056	120.934	174.962	55.782
148Asp	9.690	117.741	177.075	55.121
149Lys	7.227	119.511	173.355	53.449
150Asp	7.570	116.703	176.941	54.238
151Ala	8.781	120.352	179.637	51.735
152Asp	8.068	115.966	179.525	55.521

Appendix I

153Tyr	8.333	122.494	177.949	57.160
154Val	8.990	124.035	178.012	61.813
155Arg	9.375	122.496	181.029	68.705
156Thr	8.618	116.825	177.593	60.369
157Phe	9.716	126.936	178.581	66.504
158Phe	9.106	115.290	179.908	62.052
159Gln	8.230	122.545	180.528	62.700
160Arg	7.770	124.177	175.075	59.284
161Leu	6.959	112.787	175.999	58.101
162Asn	8.132	113.902	174.313	53.691
163Met	7.914	114.555	176.800	53.620
164Asn	10.276	127.035	174.723	52.952
165Asp	8.268	115.102	177.114	50.706
166Arg	7.801	118.360	177.455	57.800
167Glu	8.082	117.909	178.564	60.226
168Val	8.737	118.871	177.668	59.457
169Val	8.249	117.743	180.011	67.684
170Ala	9.073	123.485	179.953	66.656
180Thr	7.974	117.076	176.886	55.624
181His	7.970	122.262	178.013	55.362
182Leu	9.429	132.722	181.180	55.292
183Lys	9.228	114.725	177.363	59.946
184Asn	8.417	116.449	176.467	57.760
186Gly	8.940	111.610	172.331	59.371
187Tyr	8.527	120.658	176.537	45.725
188Glu	8.229	118.738	176.492	59.022
189Gly	7.659	118.418	178.896	55.189
193Ala	9.156	122.610	177.395	43.877
194Ala	8.233	124.680	175.979	52.463
195Asn	8.031	113.829	173.486	50.924
196Asn	7.991	115.433	173.730	54.294
197Val	7.437	117.270	173.369	51.863
198Phe	9.211	128.817	172.970	60.774
199Thr	6.447	117.546	175.247	59.111
200Asn	8.470	116.346	174.963	59.820
201Glu	7.898	119.185	173.988	52.212
202Phe	6.542	116.663	178.548	58.813
205Asn	7.883	118.492	177.625	56.817
206Leu	6.941	122.553	177.512	54.226
207Leu	6.478	112.241	178.920	57.405
208Asn	7.893	114.290	176.689	55.769
209Glu	7.207	118.699	173.424	53.949
210Asp	7.770	119.701	174.974	54.544
211Trp	7.543	124.650	176.437	52.163
212Lys	8.950	124.230	173.958	55.812
213Leu	7.987	131.241	176.297	54.657
214Glu	8.851	127.401	174.265	54.439
215Lys	8.280	120.155	177.077	54.019
216Asn	8.606	123.169	178.255	55.215
217Asp	8.881	117.847	176.887	50.707
218Ala	8.335	122.976	176.059	56.606
219Asn	8.349	113.180	173.988	51.284
220Asn	7.678	116.046	174.647	54.519
221Glu	8.525	120.327	175.478	51.565
222Gln	8.868	117.669	171.925	55.649
223Trp	8.007	123.460	174.395	53.717
224Asp	9.253	120.186	176.264	55.506
225Ser	8.518	118.702	176.777	51.694
226Lys	8.466	124.920	177.429	54.528

227Ser	7.130	112.957	172.897	57.716
230Met	8.320	123.546	171.389	51.924
228Gly	7.356	106.505	173.835	57.287
231Met	7.828	111.944	175.689	52.543
236Tyr	7.025	118.350	176.883	55.813
237Ser	7.396	113.637	175.356	59.511
238Leu	6.948	120.170	176.107	61.469
239Ile	6.642	134.111	176.947	55.692
240Gln	6.977	120.816	175.807	59.175
241Asp	7.200	123.035	173.799	55.741
243Lys	7.409	118.861	180.282	64.573
244Tyr	8.436	121.292	178.788	58.186
245Leu	8.533	122.153	178.018	57.218
246Ser	6.879	110.005	176.361	57.769
247Ile	7.058	123.492	176.452	60.992
248Val	8.121	122.048	177.956	63.825
249Lys	7.731	115.405	178.757	65.876
250Glu	6.909	120.346	179.811	59.368
251Tyr	7.349	119.729	178.435	58.732
252Ala	8.061	121.579	177.246	55.112
253Asn	6.907	112.197	174.806	53.917
254Asp	7.413	120.344	174.110	53.525
255Gln	8.909	125.969	176.677	52.611
256Asp	7.992	118.908	178.371	59.141
257Lys	7.500	122.040	177.926	57.436
258Phe	7.259	118.618	175.920	58.216
259Phe	8.880	121.277	178.270	58.259
260Lys	7.886	119.066	179.855	59.687
261Asp	8.378	122.047	179.950	59.260
262Phe	9.903	123.904	176.625	56.848
263Ser	8.410	113.287	176.709	62.960
264Lys	7.283	117.637	179.608	61.465
265Ala	8.311	123.506	179.099	58.646
266Phe	9.756	120.906	177.838	55.161
267Glu	8.181	118.043	177.331	60.283
268Lys	7.807	119.180	179.001	59.813
269Leu	8.681	120.318	179.842	59.372
271Glu	7.728	120.072	176.194	56.258
272Asn	7.254	123.505	176.478	55.167
273Gly	8.340	110.204	174.581	53.042
274Ile	7.582	122.040	175.209	45.349
275Thr	8.719	125.179	172.619	60.975
276Phe	9.111	129.452	173.731	61.483
278Lys	8.777	122.052	176.805	61.915
279Asp	8.150	115.264	175.817	58.433
280Ala	7.491	124.371	175.356	52.559
282Ser	8.318	119.385	171.205	62.112
284Phe	9.028	124.697	174.918	62.965
285Ile	7.888	120.528	176.766	54.220
286Phe	9.969	130.419	177.230	58.864
287Lys	8.753	124.062	176.568	58.906
288Thr	8.595	109.759	179.205	54.845
288Thr	8.577	109.750	180.548	54.803
289Leu	9.574	122.823	180.946	59.777
290Glu	8.896	120.433	178.649	59.002
291Glu	8.028	120.655	175.807	58.959
292Gln	7.573	116.530	174.120	58.370
293Gly	7.952	110.476	182.344	55.714
294Leu	7.704	126.333	175.356	45.852

TABLE A2: ^1HN and ^{15}NH chemical shifts (in ppm) of the NMR resonance assignments backbone nuclei of the minor form observed in the spectra of yeast CcP.

CcP residue	^1HN chemical shift	^{15}NH chemical shift
4Leu	8.211	127.066
5Val	8.172	124.518
6His	8.702	124.835
8Ala	8.470	135.095
9Ser	9.434	121.204
10Val	8.549	128.430
12Lys	8.270	124.293
13Gly	8.839	113.130
14Arg	7.581	118.710
15Ser	9.453	119.588
16Tyr	8.963	122.010
17Glu	8.865	116.833
19Phe	7.397	116.022
24Asn	8.718	117.548
27Ala	9.111	122.083
28Leu	9.285	121.490
29Lys	7.540	123.156
30Leu	8.685	120.579
31Arg	7.431	117.357
32Glu	7.860	118.102
33Asp	8.647	122.926
34Asp	8.298	119.478
35Glu	8.319	117.726
36Tyr	6.837	121.258
37Asp	8.102	128.056
38Asn	8.460	114.887
39Tyr	7.445	108.979
40Ile	7.905	124.193
41Gly	8.661	107.297
42Tyr	9.531	115.800
43Gly	9.654	109.428
53Thr	8.182	117.988
57Trp	6.869	121.758
58Asp	7.378	125.736
59Lys	6.814	122.668
60His	8.845	118.928
61Asp	6.851	114.212
63Thr	7.612	108.327
64Gly	8.373	105.350
65Gly	7.891	107.436
66Ser	8.573	111.913
67Tyr	8.429	122.463
68Gly	8.383	98.886
69Gly	7.638	109.682
70Thr	7.611	112.193
72Arg	6.613	107.937
73Phe	8.029	123.413
74Lys	8.659	121.901
75Lys	8.419	117.397
76Glu	6.386	118.542
78Asn	7.591	112.371
79Asp	6.803	123.552
84Gly	8.314	114.340
86Gln	9.340	121.521
87Asn	7.407	116.646

88Gly	7.542	107.508
89Phe	7.868	122.885
90Lys	8.510	117.917
91Phe	7.144	121.056
92Leu	7.212	114.079
93Glu	7.653	122.892
95Ile	6.625	119.056
96His	8.218	120.927
97Lys	7.589	115.204
98Glu	7.070	119.255
99Phe	7.365	116.669
104Ser	10.055	120.220
105Gly	9.414	108.839
106Asp	7.578	124.470
107Leu	7.837	122.014
108Phe	8.630	116.167
110Leu	8.712	127.342
111Gly	8.802	109.238
112Gly	7.221	105.345
115Ala	7.902	121.118
117Gln	8.136	114.951
118Glu	9.261	119.881
119Met	7.755	120.298
120Gln	7.622	108.941
121Gly	8.223	105.876
123Lys	7.858	112.491
124Ile	8.762	125.527
126Trp	9.166	124.636
127Arg	7.897	126.503
132Asp	8.097	123.959
133Thr	8.629	116.694
135Glu	8.604	124.126
136Asp	8.434	117.766
137Thr	7.901	110.371
138Thr	7.450	123.778
140Asp	8.182	119.882
141Asn	8.511	117.257
142Gly	7.955	110.458
143Arg	8.506	119.777
144Leu	7.349	120.464
148Asp	9.347	117.003
149Lys	6.987	118.844
150Asp	7.456	116.463
151Ala	8.718	120.058
152Asp	8.011	115.604
153Tyr	8.234	122.015
154Val	8.692	123.456
155Arg	9.199	121.898
156Thr	8.502	116.480
157Phe	9.607	126.572
158Phe	8.899	114.601
159Gln	8.166	122.269
160Arg	7.657	123.841
161Leu	6.908	112.442
162Asn	8.121	113.612
163Met	7.834	114.150
164Asn	10.251	126.723
165Asp	8.232	114.720
166Arg	7.770	117.957

Appendix I

167Glu	8.029	117.495
168Val	8.643	118.294
169Val	8.187	117.195
170Ala	9.008	122.951
182Leu	9.280	132.165
183Lys	9.135	114.263
184Asn	8.377	116.559
186Gly	8.798	111.141
188Glu	8.097	118.202
189Gly	7.481	118.132
193Ala	9.311	122.449
194Ala	8.314	124.324
195Asn	8.099	113.606
196Asn	7.985	115.125
197Val	7.379	116.760
198Phe	9.168	128.587
199Thr	6.577	117.305
200Asn	8.565	116.143
201Glu	8.033	118.977
205Asn	8.186	118.219
206Leu	7.309	122.600
207Leu	6.722	112.074
208Asn	8.077	114.145
209Glu	7.379	118.506
210Asp	7.898	119.545
211Trp	7.665	124.458
212Lys	9.053	124.008
213Leu	8.037	131.047
214Glu	8.905	127.114
215Lys	8.317	119.873
216Asn	8.679	122.920
217Asp	8.854	117.529
218Ala	8.303	122.628
219Asn	8.332	112.836
220Asn	7.646	115.692
221Glu	8.466	119.952
222Gln	8.861	117.545
223Trp	8.199	123.339
224Asp	9.347	119.903
225Ser	8.663	118.555
226Lys	8.545	124.710
227Ser	7.238	112.725
228Gly	7.449	106.253
230Met	8.585	123.540
231Met	8.079	112.104
236Tyr	7.206	118.402
237Ser	7.476	113.355
238Leu	7.145	120.259
239Ile	6.814	101.965
240Gln	7.091	120.589
241Asp	7.256	122.808
243Lys	7.487	118.634
244Tyr	8.470	120.999
245Leu	8.608	121.900
246Ser	6.934	109.713
247Ile	7.136	123.288
248Val	8.257	121.888
249Lys	7.844	115.172
250Glu	7.000	120.111

251Tyr	7.461	119.509
252Ala	8.204	121.398
253Asn	7.007	111.966
254Asp	7.389	120.059
255Gln	8.971	125.685
259Phe	8.911	120.974
261Asp	8.405	121.751
262Phe	9.889	123.586
263Ser	8.360	112.825
264Lys	7.250	117.268
265Ala	8.238	123.148
266Phe	9.654	120.560
267Glu	8.130	117.657
268Lys	7.748	118.839
269Leu	8.627	120.087
271Glu	7.758	119.778
272Asn	7.286	123.222
273Gly	8.373	109.899
274Ile	7.572	121.747
276Phe	9.143	129.166
278Lys	8.807	121.751
279Asp	8.183	114.975
280Ala	7.524	124.081
282Ser	8.349	119.091
284Phe	9.061	124.399
285Ile	7.902	120.231
286Phe	9.972	130.115
287Lys	8.774	123.748
288Thr	8.604	109.466
289Leu	9.587	122.516
290Glu	8.921	120.140
291Glu	8.048	120.344
294Leu	7.726	126.043

ENGLISH SUMMARY

The traditional view of protein complex formation occurring via a single step, with the proteins proceeding directly from the free to bound state, has become outdated in recent decades. In reality, most proteins and their complexes are flexible and dynamic in solution, occupying several conformations over time. Therefore, complex formation can be better thought of as following a trajectory along which a loosely associated, weakly interacting encounter complex acts to pre-orient the binding partners before they proceed to the final stereo-specific state. Both the low energy stereo-specific state and the higher energy conformations present in the encounter complex have been shown to play a crucial role in molecular recognition. However, while the stereo-specific complex is often easy to isolate and study, the encounter complex conformations comprise only a fraction of the complex and are often lowly populated, short-lived and cannot be isolated, which makes them essentially invisible to conventional structural biology techniques.

Encounter complexes often play a large role for complexes that must balance a biological requirement for a high turnover rate with the necessity of forming a specific interaction. This is particularly the case for electron transfer complexes, such as the complex between yeast cytochrome c (Cc) and cytochrome c peroxidase (CcP). The Cc-CcP is a well-studied prototypical electron transfer complex and serves as a model for the more complex reactions that occur in the respiratory electron transport chain. In the cell, two Cc molecules deliver two electrons to CcP, which uses them to reduce potentially toxic hydrogen peroxide to harmless water. Previous studies in our group, also using paramagnetic nuclear magnetic resonance (pNMR) spectroscopy, have shown that the Cc-CcP complex exists as a mixture of states in solution with approximately 70% of the population in the stereo-specific state and the remaining 30% in a highly dynamic encounter state (30:70 model). This makes the Cc-CcP complex an ideal model for studying the encounter complex.

The work described in this thesis focuses on the use of pNMR to study the invisible states of dynamic (transient) protein complexes, using the Cc-CcP complex as a model. pNMR has proven to be an extremely powerful technique for studying “invisible” states. It relies on the magnetic effects generated by an unpaired electron within a paramagnetic centre that disturb the local magnetic field experienced by nearby nuclei. This results in measurable changes in the NMR signals from which distance and orientation information for protein structure modelling can be extracted. These effects include an increase in the relaxation rate,

called paramagnetic relaxation enhancement (PRE), changes in the chemical shifts, called pseudocontact shifts (PCS), or a weak alignment with the external magnetic field, resulting in residual dipolar couplings (RDC).

Prior pNMR studies on the Cc-CcP complex were done by observing the paramagnetic effects in the spectra of Cc. In order to study such effects on CcP, the backbone amide resonance assignment was first required. A new expression construct and an updated purification protocol for isotopically labelled CcP resulted in very high yields of the triple labelled protein required for the 3D NMR assignment experiments. However, during these long experiments, each spanning several days, a second form of CcP appeared in the spectra. CcP can exist in several states depending what ligand is bound to the iron ion located within the protein's haem group. These states have different spin and coordination states but all of them are paramagnetic. By combining several analytical techniques, it was possible to identify the second form as fluoride bound CcP. This form is only present in aged samples (several days old) and does not appear in fresh samples. Unfortunately, to date the source of the fluoride could not be identified. It was also found that the resting state of CcP contains a hexacoordinate, high-spin haem coordinated to water. The paramagnetic properties and the NMR assignments were successfully obtained for both states.

With the backbone assignment of CcP complete, it was then possible to study the encounter complex from the view-point of CcP; previous studies had only looked at Cc. Spin labels (SLs) containing an unpaired electron in the form of a nitroxide radical were attached at two positions on Cc, near the stereo-specific binding interface, and the resulting PRE effects were measured in the spectra of CcP. When the new data set was compared to the existing model, the fit was good, although several novel interactions were identified as well, indicating deficiencies in the ability of the 30:70 model to completely describe the behaviour of the Cc-CcP complex in solution. This highlighted the importance of obtaining as many independent data sets as possible for achieving the most accurate description of an encounter complex.

Some of the discrepancies between the PRE data and the 30:70 model were ascribed to an underestimation of the SL mobility. SLs are attached to proteins via a single disulphide bridge. This requires only a single cysteine on the surface of the protein for attachment but also allows free rotation of the linker. This inherent flexibility of the SL label causes averaging in the NMR signal and if not corrected for during analysis, which is very difficult to do, this results in errors in the distance restraints. In order to determine the most favourable SL orientations, intramolecular PRE were measured for SLs attached on the surface of CcP. Once the most favourable SL orientations had been found, these were used for intermolecular PRE

based protein docking of Cc onto CcP. It was found that optimizing the SL orientations using intramolecular PRE resulted in solutions that were closest to the stereo-specific orientation but that SL mobility remains the limiting factor for obtaining accurate solutions, especially in this case where a large proportion of the complex is in a dynamic encounter state. Furthermore, this work underscored the limitations of PRE for determining protein orientation within a complex; PRE only provides distance restraints and can be highly biased towards minor states that come very close to the paramagnetic centre.

During this work it was discovered that, surprisingly, in the absence of Cc, an ultra-weak interaction occurred between CcP molecules in solution. The PRE effect decreases by the sixth power as distance increases, making it exquisitely sensitive to lowly populated states. During the initial intramolecular PRE measurements, no Cc was added to the sample. Unexpected PRE were observed for several nuclei more than 24 Å away from the SL, which is beyond the limit for PRE from nitroxide radicals. Furthermore, these nuclei were all relatively close to the surface of CcP and their PRE were diminished upon the addition of Cc; thus, confirming that they were, in fact, intermolecular effects. The interaction occurs between the extremely negatively charged stereo-specific binding interface and a few positively charged patches on the “back” of CcP. The formation of this ultra-weak CcP-CcP complex may have implications for the interpretation of scientific results and will need to be taken into account when designing future experiments.

The existing 30:70 model matches well with previous PRE data obtained from SLs, which are highly mobile. Caged lanthanide probes (CLaNP) are double armed tags that will bind any lanthanoid ion. The lanthanoids are particularly useful for pNMR because the strength and type of effect varies depending on the metal used. This allowing for fine-tuning of the PCS or RDC and those effects can be complemented by PREs produced by gadolinium (Gd); which produces PRE over a much larger distance than do nitroxide radicals making it ideal for studying protein complexes. Lutetium (Lu) can serve as a diamagnetic control. Therefore, CLaNP probes containing holmium (Ho) and thulium (Tm), which produced intermediate and strong effects, respectively, were attached at four positions on CcP and intramolecular PCS were recorded to determine the position, size and orientation of the $\Delta\chi$ -tensors. The intermolecular PCS, RDC and PRE were then measured for Cc.

As mentioned above, PRE are not ideal for determining protein orientations because they only provide distance information. PCS provide both distance and orientation information. RDC provide only orientation information but the strength of the effect is not dependent on the distance between the nuclei being observed and the paramagnetic centre,

making RDC useful for long distances. Therefore, we aimed to generate a new model of the Cc-CcP complex that incorporates all three types of paramagnetic data. However, the new data were first compared to the previous 30:70 model. This showed that the presence of a 3+ net charge on the CLaNP-5 probe resulted in electrostatic repulsion between the probe and the positively charged Cc. This disrupted formation of the complex and resulted in an overall poor fit between the new data and the 30:70 model, particularly for the PCS and RDC data. Therefore, other available probes with a lower net charge should be used to obtain a complete set of PCS, RDC and PRE data for the complex without disrupting complex formation. Although we recognize the computational challenges that it will present, combining these complementary data sets will hopefully allow for a more detailed model of the Cc-CcP encounter complex. Furthermore, the 30:70 model assumes that only electrostatics are involved in this interaction while hydrophobics have been shown to play a role in the formation of other ET encounter complexes. A more detailed model will hopefully allow us to determine what role, if any, hydrophobic or other short range interactions play in the Cc-CcP interaction.

This work, like many previous studies, has also demonstrated the sensitivity of pNMR to lowly populated states. For this reason, PRE has been used to characterize several encounter complexes as well as to visualize several ultra-weak intra- and intermolecular interactions, including an ultra-weak CcP-CcP association. Recently, PRE data has also led to the discovery of encounter states that do not result in a fruitful stereo-specific complex (futile complexes). The secondary Cc-CcP interaction via the low-affinity binding site has also recently been visualized using PRE data and it has been suggested that this complex is not electron transfer active and therefore plays no role in the biological Cc-CcP interaction. Since PRE data has its limitations, future studies combining PRE, RDC and PCS may also allow for a highly detailed description of this potentially futile complex.

NEDERLANDSE SAMENVATTING

Het traditionele idee dat complexvorming tussen eiwitten optreedt in één stap, waarin de eiwitten rechtstreeks overgaan van een vrije in een gebonden toestand, is in de afgelopen decennia achterhaald geraakt. In de werkelijkheid zijn de meeste eiwitten en hun complexen in oplossing flexibel en dynamisch; ze nemen meerdere conformaties aan. Daarom kan complexvorming beter gezien worden als een traject, waarbij een losjes geassocieerde ontmoetingscomplex fungeert om de partners te pre-oriënteren voordat ze het uiteindelijke stereospecifieke complex vormen. Zowel de lage-energie conformatie van het stereospecifieke complex als de hogere-energie conformaties in het ontmoetingscomplex blijken een cruciale rol te spelen bij moleculaire herkenning. Het stereospecifieke complex is vaak gemakkelijk te isoleren en te bestuderen, maar het ontmoetingscomplex vormt maar een klein deel van het complex en bestaat kort en kan daarom niet geïsoleerd worden, waardoor het in wezen voor conventionele structurele-biologietechnieken onzichtbaar is.

Ontmoetingscomplexen spelen vaak een grote rol bij die complexen waarvoor de balans tussen een korte levensduur en een specifieke interactie belangrijk is. Dit geldt voornamelijk voor elektrontransportcomplexen, zoals het complex tussen cytochroom c (Cc) en cytochroom c peroxidase (CCP) uit gist. Het Cc-CcP complex is een goed bestudeerd prototype elektronoverdrachtcomplex en dient als voorbeeld voor de complexere reacties die optreden in elektrontransportketens. In de cel leveren twee Cc moleculen elk een elektron aan CcP die gebruikt worden om potentieel giftig waterstofperoxide te reduceren tot onschadelijk water. Voorgaande studies binnen onze groep, die ook met behulp van paramagnetische kernspinresonantie (pNMR) spectroscopie zijn uitgevoerd, hebben aangetoond dat het Cc-CcP complex als een mengsel van toestanden in oplossing voorkomt, met ongeveer 70% in de stereospecifieke toestand en de resterende 30% in een zeer dynamisch ontmoetingscomplex (30:70 model). Dit maakt het Cc-CcP complex een ideaal model voor het bestuderen van het ontmoetingscomplex.

Het werk dat wordt beschreven in dit proefschrift richt zich op het gebruik van pNMR om de onzichtbare toestanden van dynamische (kortstondige) eiwitcomplexen te bestuderen, met behulp van het Cc-CCP-complex als model. pNMR heeft zich bewezen als zeer krachtige techniek voor het bestuderen van "onzichtbare" toestanden. Het maakt gebruik van de magnetische effecten die door een ongepaard elektron in een paramagnetisch centrum worden gegenereerd. Deze effecten verstoren het lokale magnetische veld van de kernen in de

omgeving van het centrum. Dit leidt tot meetbare veranderingen in de NMR signalen, waaruit informatie over afstanden en oriëntaties voor modellering van eiwitstructuren kan worden gehaald. Deze effecten zijn een toename van de relaxatiesnelheid, genaamd paramagnetische relaxatieversterking (paramagnetic relaxation enhancement, PRE), veranderingen in de chemische verschuivingen, genaamd pseudocontact verschuivingen (pseudocontact shifts, PCS) of een zwakke oriëntatie van het eiwit in het externe magnetische veld, wat leidt tot residuële dipolaire koppelingen (residual dipolar coupling, RDC).

Bij eerdere pNMR studies aan het Cc-CcP complex zijn de paramagnetische effecten in de spectra van Cc geobserveerd. Om zulke effecten bij CcP bestuderen, was de toekenning van de resonanties van de amides nodig. Een nieuw expressieconstruct en een verbeterd zuiveringsprotocol voor isotoopverrijkt CcP resulteerde in zeer hoge opbrengsten van het drievoudig isotoopverrijkte eiwit, wat nodig was voor de 3D NMR-experimenten voor toekenning. Echter, tijdens deze lange experimenten, die elk meerdere dagen duurden, verscheen een tweede vorm van CcP in de spectra. CcP kan in verschillende vormen bestaan, afhankelijk van het zesde ligand van het ijzeratoom binnen de heemgroep. Deze vormen hebben verschillende spin- en coördinatietoestanden, maar ze zijn allemaal paramagnetisch. Door het combineren van verschillende analytische technieken was het mogelijk om de nieuwe toestand te identificeren als waterstoffluoride-gebonden CcP. Deze vorm is alleen aanwezig in verouderde monsters (enkele dagen oud) en komt niet voor in verse monsters. Helaas kon de bron van de fluoride niet worden geïdentificeerd. Er is ook gebleken dat de rusttoestand van CcP een hexacoördinate heem bevat, een watergecoördineerde hoog-spin heem. De paramagnetische eigenschappen en de NMR-toekenningen zijn voor beide toestanden met succes bepaald.

Na de voltooiing van de toekenning van de amide resonanties van CcP was het mogelijk het ontmoetingscomplex vanuit het 'gezichtspunt' van CcP te bestuderen; eerdere studies hadden alleen vanuit Cc gekeken. Spin labels (SL) met een ongepaard elektron in de vorm van een zuurstofradicaal werden op twee posities op Cc, dicht bij het stereospecifieke bindingsvlak, bevestigd en de resulterende PRE werden in de spectra van CcP gemeten. De data bleken goed overeen te komen met de voorspelde waarden uit het 30:70 model, hoewel er ook enkele nieuwe interacties werden geïdentificeerd. Dit wijst op tekortkomingen van het 30:70 model om het Cc-CcP complex in oplossing volledig te beschrijven. Het is dus van belang zoveel mogelijk onafhankelijke datasets te verkrijgen om het ontmoetingscomplex zo goed mogelijk te beschrijven.

Een deel van de verschillen tussen de PRE-data en het 30:70 model wordt toegeschreven aan een onderschatting van de mobiliteit van het SL. SLs worden via een enkele disulfidebrug aan eiwitten bevestigd. Dit vereist slechts één cysteine aan het eiwitoppervlak voor bevestiging, maar maakt wel vrije rotatie van de SL-arm mogelijk. De inherente flexibiliteit van het SL veroorzaakt middeling van het NMR signaal en heeft fouten in de afstandsbepalingen tot gevolg, indien ervoor niet gecorrigeerd wordt tijdens de analyse (wat zeer moeilijk is). Om de meest gunstige oriëntaties van het SL te bepalen werden intramoleculaire PRE, van SLs bevestigd aan het oppervlak van CcP gemeten. Nadat de gunstigste SL-oriëntaties waren gevonden, werden deze gebruikt voor dockingsberekeningen van Cc op CcP op grond van intermoleculaire PRE. Het blijkt dat het optimaliseren van de oriëntaties van het SL met intramoleculaire PRE leidt tot complexen die dichter bij de verwachte oriëntatie van het stereospecifiek complex liggen. De mobiliteit van het SL blijft de beperkende factor voor het verkrijgen van een nauwkeurig resultaat, zeker in dit geval, waar een deel van het complex zich in een dynamische ontmoetingscomplex bevindt. Bovendien onderstreept dit werk de beperkingen van PRE voor het bepalen van de oriëntatie van eiwitten in een complex; PRE levert alleen afstanden op en die kunnen in grote mate beïnvloed zijn door toestanden met een lage populatie maar korte afstanden tot het paramagnetische centrum.

Tijdens dit werk werd, verrassenderwijs, ontdekt dat, bij afwezigheid van Cc, een uiterst zwakke interactie tussen CcP moleculen in oplossing plaatsvond. De PRE neemt met de afstand tot de zesde macht af wanneer deze toeneemt, waardoor de PRE uiterst gevoelig is voor de aanwezigheid van een vorm met lage populatie. Tijdens de eerste intramoleculaire PRE metingen werd geen Cc toegevoegd aan het monster. Onverwacht werden toen PRE waargenomen voor enkele kernen met een afstand tot het SL van meer dan 24 Å, wat voorbij de limiet voor PRE van nitroxide-radicalen is. Bovendien bevinden deze kernen zich relatief dicht bij het oppervlak van CcP, en de PRE namen af na de toevoeging van Cc, wat bevestigt dat het inderdaad intermoleculaire effecten zijn. De interactie vindt plaats tussen het extreem negatief geladen stereospecifieke bindingsvlak en een paar positief geladen delen op de "achterzijde" van CcP. De vorming van dit ultra-zwakke CcP-CcP complex kan gevolgen hebben voor de interpretatie van wetenschappelijke resultaten en er moet rekening mee gehouden worden bij het ontwerpen van toekomstige experimenten.

Het bestaande 30:70 model werd met behulp van SLs ontwikkeld, die aanzienlijke mobiliteit vertonen. Gekooide lanthanidesondes (caged lanthanoid probes, CLaNP) zijn tweearmige sondes die elk lanthanide-ion kunnen binden. De lanthaniden zijn nuttig voor pNMR omdat het type paramagnetische effect en de sterkte ervan varieert, afhankelijk van het

gekozen metaal. Dit maakt het mogelijk het bereik van de PCS of RDC af te stellen en deze effecten kunnen weer worden aangevuld door PRE veroorzaakt door gadolinium (Gd); Gd veroorzaakt PRE over een veel grotere afstand dan nitroxide radicalen, waardoor het ideaal is om eiwitcomplexen te bestuderen. Lutethium fungeert als een diamagnetische controle. Daarom zijn CLaNP probes met holmium (Ho) en thulium (Tm) gebruikt, welke respectievelijk middelmatige- en sterke effecten produceren. Ze zijn op vier posities op CcP bevestigd en intramoleculaire PCS zijn opgenomen om de positie, grootte en oriëntatie van de $\Delta\chi$ -tensoren te bepalen. De intermoleculaire PCS, RDC en PRE zijn vervolgens gemeten voor Cc.

Zoals hierboven vermeld, zijn PRE niet ideaal voor het bepalen van eiwtoriëntaties, omdat ze alleen afstands-informatie kan bieden. PCS geeft informatie over zowel afstand als oriëntatie. RDC geeft alleen informatie over oriëntatie van het eiwit, maar de sterkte van het effect is niet afhankelijk van afstand tussen de kern en het paramagnetisch centrum, waardoor de RDC nuttig is voor lange afstanden. Het doel van dit deelproject was een nieuw model voor het Cc-CcP ontmoetingscomplex te genereren op grond al deze typen paramagnetische data. Echter, de experimenten toonden aan dat de aanwezigheid van een 3+ lading van de CLaNP-5 sonde resulteerde in elektrostatistische afstoting tussen de sonde en het positief geladen Cc. Dit verstoorde helaas de vorming van een complex en resulteerde in een slechte overeenkomst tussen de nieuwe data en het 30:70 model, vooral voor de PCS en RDC data. Om deze reden dienen andere beschikbare sondes met een lagere lading gebruikt te worden om een complete set PCS, RDC en PRE data van het complex te verkrijgen zonder verstoring van de interactie. Het combineren van deze complementaire datasets levert hopelijk een meer gedetailleerd model op van het Cc-CcP ontmoetingscomplex, hoewel we de computationele uitdaging die dit met zich meebrengt erkennen. Daarnaast gaat het 30:70 model er vanuit dat alleen elektrostatistische interacties betrokken zijn bij deze complexvorming, terwijl aangetoond is dat hydrofobe interacties een rol kunnen spelen bij de vorming van andere ET ontmoetingscomplexen. Een gedetailleerder model zal het hopelijk mogelijk maken te bepalen of hydrofobe of andere korte-afstand interacties een rol spelen in de Cc-CcP interactie.

Dit werk heeft, net als vele andere studies, de gevoeligheid van pNMR voor vormen met lage populatie aangetoond. Om deze reden worden PRE niet alleen gebruikt voor de karakterisatie van diverse ontmoetingscomplexen, maar ook voor de visualisatie van enkele zeer zwakke intra- en intermoleculaire interacties, waaronder een zeer zwakke CcP-CcP associatie. Recentelijk heeft PRE data ook geleid tot de ontdekking van ontmoetingscomplexen die niet resulteren in een stereospecifiek complex (futile complexes). Er bestaat een secundaire

Nederlandse samenvatting

Cc-CcP interactie die via de bindingsplaats met lage affiniteit plaatsvindt, en deze is recent gevisualiseerd met behulp van PRE data en er is gesuggereerd dat dit complex geen elektronoverdracht vertoont en daarom geen rol speelt in de biologische Cc-CcP interactie. Aangezien PRE data beperkingen hebben, kunnen toekomstige onderzoeken waarin PRE, RDC en PCS data worden gecombineerd wellicht leiden tot een zeer gedetailleerde omschrijving van dit potentiële futele complex.

LIST OF PUBLICATIONS

Schilder J., Hass M.A.S., Keizers P.J.H. and Ubbink M. (2012). *Paramagnetic NMR spectroscopy and lowly populated states*. (M. G. Clore & J. Potts, Eds.). Recent Developments in Biomolecular NMR: Royal Society of Chemistry. 130-150.

Schilder J. and Ubbink M. (2013). *Formation of transient protein complexes*. *Curr Opin Struct Biol* **23**(6): 911-918.

Schilder J., Löhr F., Schwalbe H. and Ubbink M. (2014). *The cytochrome c peroxidase and cytochrome c encounter complex: the other side of the story*. *Febs Letters* **588**(10): p. 1873-1878.

Hass M.A.S., Liu W.M., Agafonov R., Otten R., Phung L., **Schilder J.**, Kern D. and Ubbink M. (2015). *A minor conformation of a lanthanide tag on adenylate kinase characterized by paramagnetic relaxation dispersion NMR spectroscopy*. *J Biomol NMR* **61**(2): 123-136.

Schilder J., Liu W.M., Kumar P., Overhand M., Huber M. and Ubbink M. (2015). *Protein docking using an ensemble of spin labels optimized by intra-molecular paramagnetic relaxation enhancement*. *Phys Chem Chem Phys.*, DOI: 10.1039/c5cp03781f.

Schilder J., Kumar P., Florea B., Huber M. and Ubbink M. (2015). *NMR assignment and analysis of haem iron $\Delta\chi$ -tensors for two forms of yeast cytochrome c peroxidase: serendipitous discovery of time-dependent hydrogen fluoride binding. To be submitted.*

Schilder J., Liu W. and Ubbink M. (2015). *Ultra-weak self-association between cytochrome c peroxidase molecules observed by paramagnetic NMR. To be submitted.*

CURRICULUM VITAE



Jesika Schilder was born on March 24th 1986 in Whitehorse, Yukon, Canada. She began her Bachelor of Biochemistry in 2005 at the University of Victoria. She received the President's Scholarship for four years before graduating with Honours with Distinction in 2009. During her bachelor's degree, she completed three research internships: the first with Prof. Dr. Viktoria Weber at the Danube University Krems in Krems an der Donau, Austria; the second with Sir Prof. Dr. John E. Walker, FRS, at the MRC Dunn Human Nutrition unit in Cambridge, U.K.; and the third with Prof. Dr. Francis Nano at the University of Victoria, Victoria, Canada. This solidified her love for lab-work as well as her desire to move to Europe.

In September 2009, she was able to begin her Master in Chemistry at the Universiteit van Amsterdam with the help of the H.S.P. Huygens scholarship. For her final Master internship, she decided to join the Protein Chemistry group at Leiden Universiteit with Prof. Dr. Marcellus Ubbink. She enjoyed the work very much and was able to stay for her PhD. During her PhD she used paramagnetic NMR spectroscopy to study the highly dynamic electron transfer complex between cytochrome *c* and cytochrome *c* peroxidase, the results of which have been presented in this thesis.

Aside from her academic work, Jesika has become proficient in Dutch and is also an avid scuba diving instructor. She has also managed to visit almost 40 countries across the Americas, Europe, Asia and Africa. Over the next year she intends to work and travel through the Caribbean and the Indo-pacific before discovering what her future holds.