Dynamics in photosynthetic transient complexes studied by paramagnetic NMR spectroscopy

Scanu, S.

Citation

Scanu, S. (2013, October 10). *Dynamics in photosynthetic transient complexes studied by paramagnetic NMR spectroscopy*. Retrieved from https://hdl.handle.net/1887/21915

Version:	Not Applicable (or Unknown)
License:	Leiden University Non-exclusive license
Downloaded from:	https://hdl.handle.net/1887/21915

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/21915 holds various files of this Leiden University dissertation.

Author: Scanu, Sandra
Title: Dynamics in photosynthetic transient complexes studied by paramagnetic NMR spectroscopy
Issue Date: 2013-10-10

.

Dynamics in photosynthetic transient complexes studied by paramagnetic NMR spectroscopy

Sandra Scanu

Dynamics in photosynthetic transient complexes
studied by paramagnetic NMR spectroscopy

Sandra Scanu

Doctoral Thesis, Leiden University, 2013

ISBN number: 978-94-6203-444-0

© 2013, Sandra Scanu

Cover designed by Maddalena Idili: it.linkedin.com/pub/maddalena-idili/40/374/883

Printed by CPI-Wöhrmann Print Service - Zutphen

Dynamics in photosynthetic transient complexes studied by paramagnetic NMR spectroscopy

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 10 Oktober 2013
klokke 15.00 uur

door
Sandra Scanu
Geboren te Sardinië, Italië
in 1981

Promotiecommissie

Promotor:	Prof. Dr. M. Ubbink
Overige leden:	Prof. Dr. R. Boelens (Universiteit Utrecht) Prof. Dr. J. Brouwer Dr. I. Díaz-Moreno (Universidad de Sevilla) Prof. Dr. E. J. J. Groenen Prof. Dr. G. M. Ullmann (Universität Bayreuth)

The investigations described in this thesis were performed at the Protein Chemistry department of the Leiden Institute of Chemistry, Leiden University, Leiden, the Netherlands.

Financial support for the research was provided by the Netherlands Organisation for Scientific research (NWO), Chemical Sciences ECHO grant 700.57.011.

dedicated to my beloved sister
who always encourages me
to follow my aspirations

Contents

Abbreviations **7**

Chapter I **11**
Introduction

Chapter II **25**
The complex of cytochrome *f* and plastocyanin from *Nostoc sp.* PCC 7119 is highly dynamic

Chapter III **39**
Role of hydrophobic interactions in the encounter complex formation of plastocyanin and cytochrome *f* complex revealed by paramagnetic NMR spectroscopy

Chapter IV **65**
Loss of electrostatic interactions causes increase of dynamics within the plastocyanin-cytochrome *f* complex

Chapter V **87**
Concluding remarks

Nederlandse Samenvatting **93**

English Summary **97**

References **101**

Appendices **111**

List of publications **137**

Curriculum vitae **138**

Abbreviations

ET	Electron transfer
Pc	Plastocyanin
Cyt *f*	Cytochrome *f*
N	*Nostoc sp.* PCC 7119
Ph	*Phormidium laminosum*
b_6f	Cytochrome *b_6f*
PSI	Photosystem I
NMR	Nuclear magnetic resonance
HSQC	Heteronuclear single quantum coherence
CSP	Chemical shift perturbation
PCS	Pseudocontact shift
PRE	Paramagnetic relaxation enhancement
BD	Brownian dynamics
MC	Monte Carlo
MES	2-(*N*-morpholino) ethanesulfonic acid
MTS	(1-Acetoxy-2,2,5,5-tetramethyl-δ-3-pyrroline-3-methyl) methanethiosulfonate
MTSL	(1-Oxyl-2,2,5,5-tetramethyl-δ-3-pyrroline-3-methyl) methanethiosulfonate
CoM	Center of mass

“Così tra questa immensità s’annega il pensier mio:
e il naufragar m’è dolce in questo mare.”

- In such immensity my thinking drowns and it is sweet to shipwreck in this sea-

Giacomo Leopardi, L’infinito, 1818-1819.

Chapter I

Introduction

Based on the review article:
Bashir Q., Scanu S., and Ubbink M. Dynamics in electron transfer protein complexes. *FEBS J*. **278**, 1391-1400 (2011).

Transient complexes

Protein-protein complexes can be classified as static or transient, on the basis on their binding characteristics. The equilibrium dissociation constant (K_D) is given by the ratio between the dissociation rate constant (k_{off}) and the association rate constant (k_{on}).[1] The k_{off} values range over many orders of magnitude and, consequently, so does the life-time of the complex (1/ k_{off}). Static complexes can have K_D values as low as 10^{-16} M, indicating life-times that can be as long as days, in the extreme case of toxins bound to their targets.[2] An example of a static complex formed by an enzyme and its inhibitor, is the complex of TEM1-β-lactamase (TEM1) and β-lactamase inhibitor protein (BLIP).[3] Association is not very fast (k_{on}= 2.6 x 10^5 M^{-1} s^{-1}), but the low k_{off} (2 x 10^{-4} s^{-1}) ensures that the proteins dissociate slowly. Even though the thermodynamic affinity is not very high (K_D = 10^{-9}M), the enzyme is kinetically shut down. On the other hand, K_D values for transient complexes vary in the range of μM to mM,[4] resulting from a high k_{off} (≥ 10 s^{-1})[5-8] and k_{on} in the range of 10^7-10^9 M^{-1} s^{-1}.[4,9] The lifetimes are thus on the μs-ms time scale. Proteins involved in signal transduction and electron transfer pathways are examples of such complexes. In all cases, the lifetime of the complex is closely related to its function. The transmission of a signal and the transfer of an electron are fast events and a high turnover of the involved partners is required, therefore, rapid dissociation is cardinal.

High geometrical surface complementarity between the interaction partners is a prerequisite in static complex formation. A closely packed interface is formed and stabilized by a dense network of short range forces, such as hydrogen bonds and van der Waals, hydrophobic and short-range electrostatic interactions. At the same time the multitude of short-range interactions leads to high specificity. There are specific residues, which contribute most to the binding,[10] so-called "hot spots" of binding and their presence is characteristic in static complexes.[11,12] Due to the lack of crystal structures, the ubiquity of "hot spots" in transient complexes has not yet been definitively established,[13] and they have been visualised only in a couple of cases.[14,15] Transient electron transfer (ET) protein complexes do not present a high level of geometric complementarity on the binding surface and the specificity is low. Instead, the protein surfaces are often optimized for the recognition of multiple interaction partners.[16] Such promiscuity is of paramount importance for the physiological function of the proteins.[14] For example, an electron carrier in photosynthesis or mitochondrial respiration needs to transfer an electron from a donor to an acceptor protein(s), so that it has to recognize specifically at least two partners, carry out ET and quickly dissociate. Proteins involved in ET usually hold the redox centre in the proximity of a surface hydrophobic patch, which represents the interaction patch for the reaction partner(s).[17] In this way, the ET pathway from the redox centre to the protein surface is optimised to be as short as possible. In the cupredoxin family,[18,19] as in many other cases, this hydrophobic region is surrounded by polar amino acids, which assist in dissociation since they facilitate the entry of water molecules in the binding interphase.[4] Finally, in order to be biologically functional a protein-protein complex requires a fine balance of association and dissociation rates, of affinity and specificity.[20]

Due to the interest in the ET-phenomenon and associated events, and the availability of biochemical and biophysical techniques applicable to these systems, the redox partners in photosynthesis and mitochondrial respiration are, at the moment, the best understood transient complexes.

Protein-protein interactions

Protein-protein interactions represent a fundamental biological phenomenon, being the basis of essentially all cellular processes, such as signal transduction regulation, electron transfer, chemical reactions, molecular recognition in the immune system, cytoskeletal movement, and more. A thorough understanding of the ways in which proteins recognise each other and the physical forces involved is paramount in addressing many questions pertaining to the events behind the interaction processes. The formation of a protein complex is a multistep reaction.[21] Initially, proteins come in contact with each other by random collisions and form an encounter complex, which is the end point of diffusional association.[22] In the first step, proteins diffuse by thermal motions and the presence of opposite charges leads the pre-orientation of the proteins by long-range electrostatics.[23] The second step consists in the transition from the encounter (AB*) to the final complex (AB), in which short-range interactions, such as van der Waals interactions, hydrogen bonds, hydrophobic interactions and short-range electrostatic interactions, dominate (Figure 1.1). An encounter complex does not always evolve to a final complex, in many cases it actually represents a futile complex (AB**).[22]

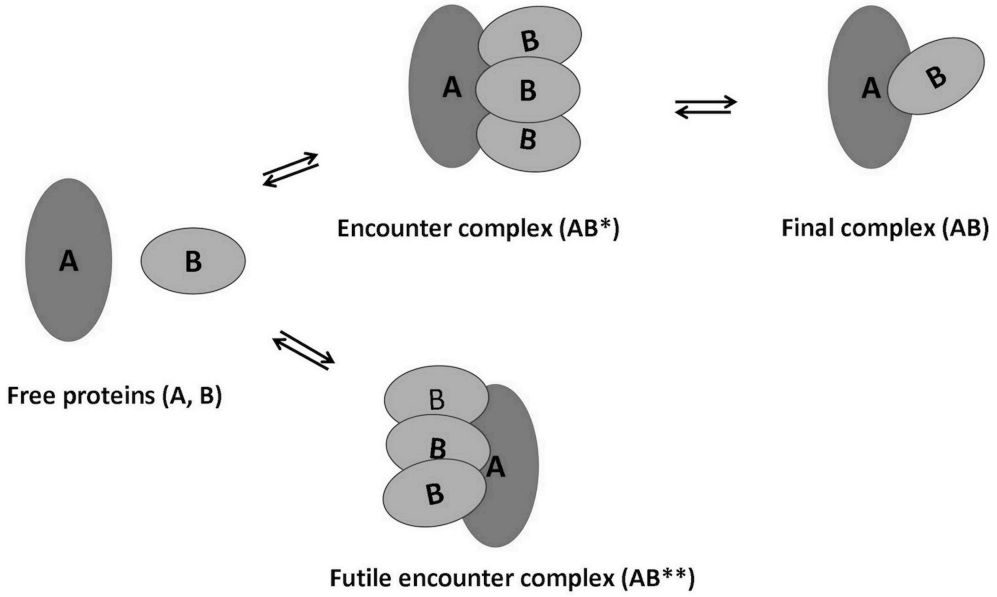

Figure 1.1. Two-step model of protein complex formation.[21] Free proteins (A and B) diffuse, and form either an encounter complex (AB*), which is in equilibrium with the final complex (AB) or a futile complex (AB**), which is in equilibrium with the free proteins.

As transient complex formation involves fast association and dissociation, proteins need to finely balance specificity and rapid turnover. When proteins meet by diffusion, they stay associated for some time, even in the absence of interactions. This is called the macrocollision. During the macrocollision, proteins have to find the target site on the counterpart surface to form the final complex. The target site of a protein represents only a small percentage of the overall surface (often <1%), so the search requires the sampling of an extended area, which takes time. The lifetime of macrocollisions is often insufficient to guarantee a productive search (futile encounter).[24] The formation of the encounter complex, consisting of an ensemble of multiple protein-protein orientations, optimises the number of productive collisions by increasing the chance of final complex formation via two mechanisms, both depending upon electrostatic forces. In the first mechanism, the encounter complex formation can decrease the dimensionality of the search. Electrostatic forces extend the lifetime of the macrocollisions, therefore the three-dimensional diffusion and search of the binding site is reduced to two-dimensional surface translations and reorientations (microcollisions). In the second mechanism, opposite charges pre-orientate the surfaces of the proteins in a way that the active sites face each other. The area to sample is thus drastically reduced and proteins can use the available time optimally to find the active site, thus increasing k_{on} up to four orders of magnitude. In this way, the formation of the encounter complex enables the proteins to orientate their interaction sites into a potentially fruitful position for interaction, via translational and rotational movements on the reciprocal surfaces. The formation of this intermediate leads to a stabilised active complex via the aforementioned short-range interactions. Consequently, the encounter complex is in equilibrium between the dissociated components and the specific state[25] and the equilibrium depends on the balance between non-specific long-range electrostatic forces and specific short-range forces. The population of the encounter state varies significantly among different complexes, dependent on the nature of the complex, so that it may correspond to a large fraction of the complex.[26-29] In highly dynamic systems, the encounter complex can form the larger population of the complexes, or the entire population is encounter complex.[30] Mutation of surface residues involved in the association of the complex, for example, those important for electrostatic interactions, can shift the equilibrium towards the specific or the encounter complex.[31,32] Moreover, such mutations can produce a fruitful or futile encounter complex.[24]

Therefore, the encounter complex must be explicitly considered in the study of protein complexes, in order to achieve a complete description of the interaction process.

Photosynthetic electron transfer proteins

The subject of this thesis is the transient complex formed by plastocyanin (Pc) and cytochrome *f* (Cyt *f*), two redox partners in oxygenic photosynthesis. Photosynthesis takes place at the thylakoid membrane of chloroplasts in plants, algae and cyanobacteria (Figure 1.2). It is a fundamental biochemical process,

through which light energy is converted into chemical energy, leading to the production of ATP and NADPH molecules. These compounds provide the energy and reducing power for biosynthetic pathways, including carbon assimilation.[33] Dioxygen is produced as a waste product. The photosynthetic machinery consists of a chain of membrane-bound protein complexes and mobile partners, which accomplish the long-range electron transfer process. When light hits photosystem II (directly or indirectly via the light harvesting proteins), the water splitting reaction is triggered and two water molecules are consecutively oxidized to generate four electrons, one oxygen molecule and four protons. Protons contribute to the membrane proton gradient, which is the driving force for the ATP production. The electrons are transferred through the membrane from photosystem II (PSII) to the cytochrome b_6f complex (b_6f) with the help of the lipid-soluble compound plastoquinone (PQ). Plastocyanin (Pc) is a soluble mobile electron carrier, which shuttles electrons from b_6f complex to photosystem I (PSI). Finally, electrons are transferred to ferredoxin-NADP reductase (FNR) via ferrredoxin (Fd) for the production of NADPH. The rapid ET in the photosynthetic chain requires a fine balance between the specificity of the interactions among different partners and the rapid turnover of the complexes, which is ensured by the transient interactions.

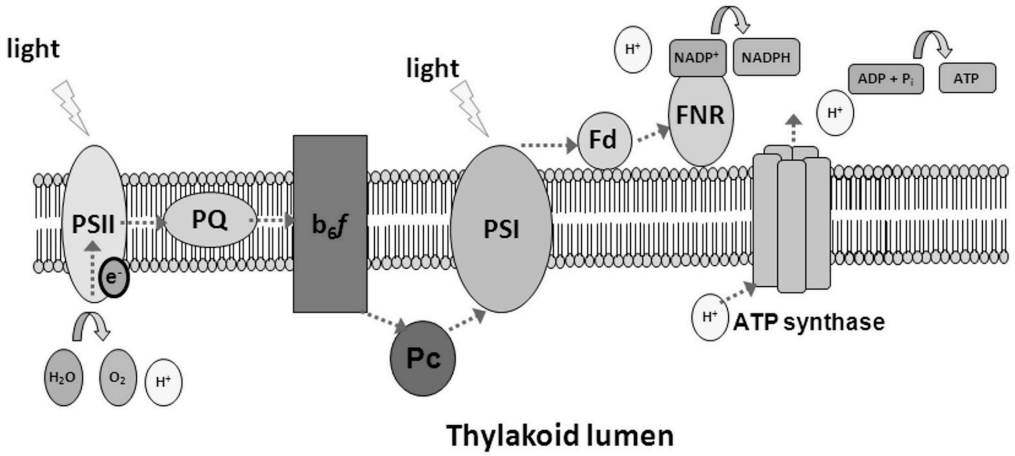

Figure 1.2. Schematic representation of electron transfer in the photosynthetic process.

In some cyanobacteria and green algae, the copper protein Pc can be substituted by the haem protein cytochrome c_6 (cyt c_6), which has the same function. Although the two proteins have no structural similarities, their interaction surfaces seem to be evolutionary correlated.[34] It has been proposed that the alternative expression of the genes coding for the proteins depends on the relative environmental abundance of copper and iron. In this way cyanobacteria and green algae may adapt to chemically different environments. In recent years, various new c-type cytochromes have been identified in plants (cyt c_{6A}), cyanobacteria (cyt c_{6B}) and algae (cyt c_{6C}).[35] They have structural features similar to cyt c_6 but their function has not been yet elucidated, though their low haem mid-point redox potential (E_m)[36] excludes a possible functional analogy with cyt c_6 and Pc.[37] The possibility of inter-exchange between Pc and cyt c_6 within the

photosynthetic chain underlines the importance of the transient and dynamic nature of the complexes formed by Pc and cyt c_6 with both b_6f complex and PSI.[34]

Plastocyanin

Pc was first isolated from the green alga *Chlorella ellipsoidea*[38] and then identified as electron carrier from Cyt *f* of the b_6f complex to the chlorophyll pair P_{700}^+ of Photo-system I (PSI).[39] It is a small, type I blue copper protein,[40] with a maximum UV absorption at 595 nm. The metal is coordinated in a trigonal pyramidal manner by the N_δ atoms of two histidines, one S_γ of a cysteine and one S_δ of a methionine and this motif is highly conserved among the different species (Figure 1.3).[41]

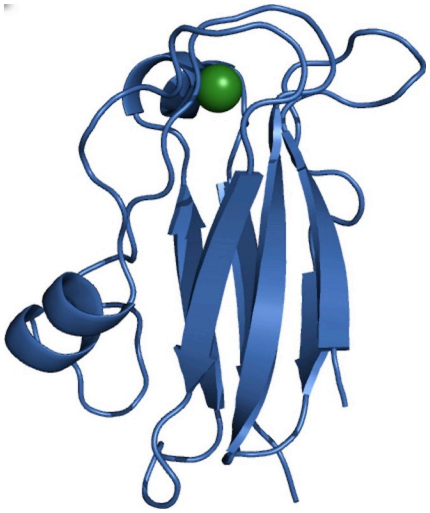

Figure 1.3. Plastocyanin from the cyanobacterium *Nostoc* (PDB entry 2GIM[42]). The backbone is represented as a blue ribbon. The copper ion is shown as a cyan sphere. This image and others of molecular structures were made with Discovery Studio Visualizer 2.5 (Accelrys®).

More than 100 genes in GenBank(www.ncbi.nlm.nih.gov/genbank) are clearly labeled as *pet*E-plastocyanin, from as many sources. Among these organisms the amino acids chain length of the mature protein varies from 97 to 105, from which the longer chains belong to cyanobacteria. Pcs have a molecular weight of 11 kDa and a highly conserved rigid β-sandwich structure[43] composed of seven β-strands and one irregular strand containing a small helix, mainly found in cyanobacteria. The metal ion and the proximal histidine ligand are surrounded by 25 amino acids that form an important region for electron transfer and protein-protein interactions, named site 1 or the hydrophobic patch,[44] which contains 16 hydrophobic residues and for which some small dynamics were observed.[45] The hydrophobic patch is present in both eukaryotic and prokaryotic Pcs, being a more extensive feature in the latter.[18] In fact, the hydrophobic patches in cyanobacterial Pcs comprise residues with long aliphatic chains, such as leucines and methionines, which make the overall region more expanded than in plant Pc. Another important feature of this protein is the presence of a ridge of charged amino acids at the side of the reaction centre, known as site 2. The nature of this site varies between eukarya and cyanobacteria. It is involved in the interaction with redox partners and influences the final orientation of the protein within the complexes it forms.[32,46-50] In plants, the acidic character of the site 2 surface exposed residues creates a negatively charged area, which interacts with a corresponding positively charged region on the small domain of Cyt *f*. In cyanobacteria, the charges are reversed with a predominance of basic exposed residues at site 2 in Pc and negatively charged amino acids in Cyt *f*.

Cytochrome *f*

Cyt *f* is the largest component of the b_6f complex.[51,52] It is embedded in the thylakoid membrane by a single C-terminal transmembrane α-helix, while the soluble part is exposed in the intraluminal space where the interaction with Pc occurs. Soluble Cyt *f*, being 254 amino acids, was first crystallized from turnip leaves (*Brassica rapa*)[53] and the structure revealed that it consists of a 28 kDa β-sheet protein with an elongated shape organized in two domains, named small and large domains. It belongs to the *c*-type cytochromes family because the haem is covalently attached to two cysteines in the conserved motif CXXCH of the protein through thioether bridges. The fifth ligand of the iron is an axial histidine and the sixth ligand of the iron is the N-terminus of the protein (Y1).[53] The presence of the haem provides the characteristic absorption profile with a Soret band at 420 nm and a α-band with a maximum at about 556 nm, in the reduced form of the protein. The haem iron in the oxidized state represents a resident paramagnetic source, that can be used for NMR studies as it gives rise to intra- and inter-molecular pseudocontact shifts, from which distance information from the iron to the observed protons can be obtained.[46] The *c*-type haem is located in the large domain, close to the hinge region with the small domain (Figure 1.4).

Figure 1.4. The structure of the soluble part of Cyt *f*, derived from the b_6f complex crystal structure from the cyanobacterium *Nostoc* (PDB entry 2ZT9). The backbone is shown as red ribbons, the haem in yellow stick representation and the iron as a grey sphere.

The porphyrin ring is located underneath a hydrophobic patch, which participates in the reaction with Pc. The contact between the hydrophobic surface on Cyt *f* and the corresponding one in Pc creates a hydrophobic interface for the ET. Cyt *f* is overall very negative, but in the small domain a charged region is present, with a composition that varies between different species, but generally is of a charge that is opposite to Pc. Therefore, it is positively charged in plants and negatively charged in cyanobacteria. The surface charge complementarity between the small domain area on Cyt *f* and site 2 on Pc plays an important role in the final orientation of the proteins with respect to each other. Hence, the architecture of the complex can be distinguished as "head on", as in the cyanobacterium *Phormidium laminosum*,[54] where Pc participates in the interaction with only the hydrophobic patch due to the lack of basic amino acids at site 2, and "side on", in plants[46-48] and in the cyanobacteria *Nostoc*[49,50] and *Prochlorothrix hollandica*,[32] where electrostatic attractions also

occur between the aforementioned sites, orientating Pc toward the small domain of Cyt *f*.

Pc-Cyt *f* complex

The Pc-Cyt *f* complex represents an excellent and challenging system for studying the influence and the contribution of different forces in ET, binding kinetics and in the structure of the complex.[55] In fact, even though the three dimensional structure of the interaction partners is highly conserved, the surface properties vary significantly in the different species.
In plants and green algae, Cyt *f* presents a ridge of positively charged amino acids, mainly composed of lysines. The electrostatic features within the complex have been the focus of extensive studies, due to the role they play in the association and in the function of the complex itself. The ET pathway from Cyt *f* into Pc was initially debated. Two main routes were proposed, via H87 (numbering of spinach Pc), which is a direct ligand for the Cu, and via Y83, in which case the electron tunneling is enhanced by a passage through C84 (which also coordinates the Cu). The ET pathway from Cyt *f* into Pc was initially debated. Two main routes were proposed, via H87 (numbering of spinach Pc), which is a direct ligand for the Cu, and via Y83, in which case the electron tunneling is enhanced by a passage through C84 (which also coordinates the Cu). The latter pathway indicated a cation-π interaction between K65 on Cyt *f* and Y83 on Pc as possible mechanism for ET.[56] A Monte Carlo approach, based on the electrostatically most favorable complexes,[57] suggested the initial formation of a high binding affinity complex in a configuration, which favored the interaction between these residues followed by a rearrangement that promoted the intermolecular ET through H87 instead of Y83. The role of electrostatics was thought to enhance the formation of the encounter complex, which would give rise to the ET-active complex. Subsequent Brownian Dynamics-based docking studies supported the ET pathways between Y1 on Cyt *f* and H87.[58] The first solution structure of the complex, published in the same year, indicated that ET must occur predominantly via H87.[46]
The transient nature of the Pc-Cyt *f* complex appears to impair the co-crystallization of the two proteins because until today no crystal structures of the complex are available. However, a plant complex was investigated via NMR using pseudocontact shifts[46] and diamagnetic chemical shift perturbation analysis.[47,48] The large size of the chemical shifts observed for the residues belonging to the hydrophobic patch on Pc, defined it, in terms of specificity, as the central binding site, suggesting a specific interaction interface with at least partial exclusion of solvent molecules. In contrast, the modest size of the shifts of the amino acids in the charged areas of the interaction partners suggested a charge-charge interaction in which the residues do not undergo desolvation. The complex was shown to exist predominantly in a state with "side on" orientation (Figure 1.5A) promoted by Coulombic attractions from the negatively charged Pc and the positively charged Cyt *f*. The role of electrostatics in the orientation of the proteins matched with kinetic measurements on Pc[59] and Cyt *f* mutants.[60] Single or bulk mutations of charged residues important for the overall electrostatic

potential of the two proteins underlined the contribution of long range electrostatic forces in the formation of the encounter complex and consequently in the overall ET rate.[61]

Further NMR investigations on the side-chain interactions validated these results and enabled the deduction of the roles of electrostatic interactions in the molecular recognition step and the contribution to the binding affinity. The complex showed the same behavior even at different ionic strength and pH conditions.[62] On the other hand, *in vivo* studies on *Chlamydomonas reinhardtii*[63,64] had opposing outcomes. The ET rate was measured for several Cyt *f* mutants, which concerned specific residues of the positively charged site, and no influence of electrostatics was observed, while *in vitro* kinetics experiments showed a monotonical decrease in the oxidation of Cyt *f* with dependence on ionic strength.[65] *In vivo* the intralumen ionic strength is higher than the one used for the *in vitro* experiments, corresponding to ~350 mM,[63] a value at which electrostatic contribution is limited. Furthermore, the volume of mature thylakoids is small and long-distance Pc diffusion is not required, contrary to situation of initial chloroplasts formation, before the membranes are enclosed, when the thylakoidal space is wider. The contradiction between *in vitro* and *in vivo* studies resides in the different composition of the media where Pc and Cyt *f* were studied, and in the latter case, the concomitant presence of thylakoidal elements of diverse nature, could somehow mask the effect of electrostatics.[62]

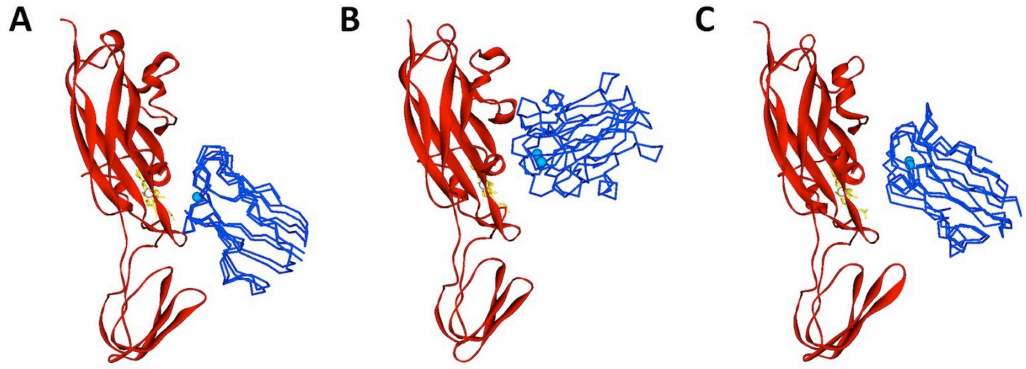

Figure 1.5. Orientations of Pc (shown as blue Cα traces) in the complex with Cyt *f* (shown as red ribbon) from different organisms: "side on" in plants (A) (PDB entry 2PCF[46]) and in the cyanobacterium *Nostoc* (C) (PDB entry 1TU2[50]), and "head on" in the cyanobacteria *Phormidium laminosum* (B).[54]

Also the complex from *Nostoc sp*. PCC7119 revealed to be electrostatically stabilised by complementary localised charges. Within this structure, Pc was found in "side-on" orientation (Figure 1.5C),[50] despite the surface charges being reversed in this organism, with a highly basic character of the copper protein (pI=8.8) compared with its eukaryotic analogues (pI=3.8 for spinach Pc). In this case, complex formation is heavily dependent on hydrophobic and electrostatic interactions in both the final state and the encounter complex.[49] The electrostatic nature of the interaction was shown by kinetic experiments on Pc[8] and Cyt *f* mutants.[66] Mutations offsetting or inverting the electrostatic charges in the charged areas and in the hydrophobic patch of both proteins highlighted the

importance of electrostatics in the complex formation and hydrophobics in ET. The ionic strength dependence of the rate constant is mainly determined by the pronounced positive charges of Pc, which are also responsible for the specificity of the interaction.[66]

Pc from *Ph. laminosum* presents a more symmetric charge distribution compared to plant and *Nostoc* orthologs. Positively and negatively charged amino acids are not found in patches but are balanced. Thus, Pc has more neutral character, with a pI= 5.0. As a consequence, the complex showed a weaker affinity (in the millimolar range, in contrast with the micromolar range in *Nostoc*) and a moderate dependence on electrostatic interactions in complex formation. Thus, electrostatic interactions only confer a minimal contribution to the final organisation of the complex as indicated by the NMR structure, [54] with Pc in a "head on" orientation (Figure 1.5B), and solely the hydrophobic patch goes to form the interaction surface with Cyt *f*. The fit of PCS data did not converge into a single structure, indicating a highly dynamic complex and suggesting that the encounter complex must have a decisive role in leading the reduction of Pc by Cyt *f*. Although the electrostatics do not appear to be important for the orientation of the proteins within the complex, kinetics[67] and theoretical[68] studies pointed to an electrostatic influence on the association process (effects on k_{on}). Also in *Ph. laminosum* Pc shows the highest dependence on electrostatics as well as a bigger influence on specificity.[69] Mutagenesis studies involving the N-terminal hydrophobic residues of Cyt *f*,[70] responsible for the shielding of the haem from the solvent, indicated that hydrophobic interactions have a role in the encounter complex formation and showed the prominent role of Cyt *f* in the fast ET rate observed at high ionic strength. The explicit treatment of the non-polar desolvation (hydrophobic) force together with the electrostatic force on the *Nostoc* and *Phormidium*[71] complexes highlighted a stronger contribution of electrostatic interactions to the overall ET rate in *Nostoc*, due to their contribution in the encounter complex formation. The role of hydrophobic interactions contributed most to the formation and stabilization in the *Phormidium* complex instead.

The complex from the cyanobacterium *Prochlorothrix hollandica* emerged to be highly dynamic as well, with a "side on" orientation.[32] The role of electrostatic interactions in the complex formation was established by observing a decrease in the affinity with increasing ionic strength, although the influence is less prominent than in plants and *Nostoc*.

The different orientations of Pc and Cyt *f* within the various complexes seem to depend on both the long-range forces involved in the intermolecular recognition and the short-range forces important for the formation of the final complex. A surprising variation in the structures and contributing interaction forces has been observed among different species. Also the degree of dynamics in the various complexes appears to differ.

NMR methods for studying transient complexes

^{15}N labeling of proteins is a well-established technique that allows to perform two-

dimensional ^{15}N-^{1}H heteronuclear single-quantum correlation experiments (HSQC). The HSQC spectrum could be considered as a two-dimensional picture of the protein. Each HSQC cross-peak correlates the chemical shift of the amide proton with the chemical shift of the attached ^{15}N amide nitrogen of a certain ^{15}N labeled amino acid residue of the protein, in which is possible to observe and quantify changes in the position and in the intensities of the HSQC resonances due to interaction with one or more partners.
Solution NMR spectroscopy is an excellent technique for the characterization and structural determination of transient complexes. In the fast exchange regime on the NMR time scale, a signal is an averaged representation of all species in solution, but it is still a powerful tool to study dynamics within a transient complex. The NMR techniques used in this thesis are diamagnetic chemical shift perturbations (CSP) and paramagnetic NMR, comprising pseudocontact shift (PCS) and paramagnetic relaxation enhancement (PRE).

Diamagnetic chemical shift perturbations

CSP occurs as a result of changes in the chemical environment of the observed nucleus. When a protein interacts with its partner, the residues involved in the interaction will usually experience a change in their chemical environment, which will cause a shift of the respective NMR resonances. CSP analyses are commonly performed on ^{15}N-^{1}H HSQC spectra of the ^{15}N labeled protein recorded in the absence and presence of increasing amounts of the unlabelled partner, or vice versa, in which the labeled protein is titrated into the unlabelled partner. In this way, it is possible to obtain a map of the amino acids involved in the interaction and visualise the binding sites. Furthermore, the analysis of chemical shifts as a function of protein concentration provides information both on stoichiometry of the complex and the binding affinity. The size of the shifts is also related to the dynamics of the complex; large chemical shift perturbations indicate well-defined, short-range interactions, whereas small ones denote high dynamics and weaker interactions.[32,72,73] Short range interactions are involved in the formation of the final complex, which is stabilized in a single orientation through a network of H-bounds, salt bridges and van der Waals interactions. Also changes in the solvation of the protein occur. Together, these interactions cause large CSPs. According to current model, the encounter complex forms by long range electrostatics interactions and the proteins are still largely solvated, so that the chemical environment at the binding sites undergoes minimal alterations and the size of perturbations is small. As a consequence, average chemical shifts values also reflect the population of the encounter state[9,30,72] and therefore can be also used as a reliable diagnostic tool for the dynamics within transient protein-protein complexes.[72-74]
The information gained from chemical shift perturbations analysis mainly has a qualitative value, therefore other NMR methods, such as paramagnetic NMR, are required for detailed structure determination.

Paramagnetic NMR

The presence of a paramagnetic centre in a molecule or in the solvent causes distance-dependent effects on the NMR signals of observed nuclei, which can provide very precise structural information. The paramagnetic source can be intrinsic to the protein, for example, the Fe^{III}-haem of Cyt *f*, or can be inserted into the protein via site-directed spin labeling (SDSL). In the latter case, a paramagnetic probe is linked to a cysteine residue that either is present on the protein or introduced by site-directed mutagenesis. A paramagnetic probe is a small molecule containing an unpaired electron on a metal chelating tag[75-79] or a stable organic radical.[80-83] The chemical-physical properties of the paramagnetic centre determine the nature of the effects on the NMR spectrum. Anisotropic electron g-tensors give rise to direction dependent effects such as residual dipolar coupling (RDC), associated to paramagnetic alignment, and PCS, together with PRE. Isotropic electron g-tensors only cause PRE.[84] PCSs arise from through-space dipolar interactions between the (time-averaged) magnetic fields of the paramagnet and the observed nucleus, which cause additive paramagnetic shifts of the observed NMR signals. The PCS effect is proportional to the inverse third power of the distance (r^{-3}) between the unpaired electron and the observed nucleus.[85] Therefore, PCSs can be measured for residues that are far from the paramagnetic centre.[86] up to 60 Å.[87] In the case of PRE, the longitudinal electron spin relaxation as well as the dipolar coupling cause nuclear relaxation,[88] which gives rise to line broadening of the nuclei in its proximity.[84] PRE shows r^{-6} distance dependency, hence the peaks corresponding to amino acids close to the paramagnetic centre will exhibit a decrease in peak height or will disappear entirely from the spectrum. PREs can be very strong and can affect nuclei up to 35 Å from the paramagnetic centre.[84]

When a complex is studied, one of the interaction partners contains the paramagnetic centre and PCSs, or PREs, are observed on the isotopically labelled counterpart monitored in the NMR spectrum.[81] This approach has been extensively used for intermolecular structure determination, taking advantage of both PCSs[46,50,54,87,89-91] and PREs.[26,77,92-97] PCS has been the method used for the determination of the structures of Pc-Cyt *f* complexes from several plants,[46,48,62] *Ph. laminosum*,[54] *Nostoc*[98] and *Pr. hollandica*.[32] It showed to be appropriate for the visualization of the complex in the highest populated state (AB in Figure 1.1). However, considering the high dynamics within this complex, as suggested by NMR CSP studies and kinetic data, PRE measurements on this complex will certainly provide a more complete picture of the overall complex. The PRE method is an exquisitely sensitive technique to detect lowly populated states of proteins orientations, provided the paramagnetic effect is much larger for the lowly populated state than the ground state.[92] This is not the case of the haem derived PCS, so minor states are not easily detected using only these shifts. The introduction of a paramagnetic source in different locations on the Cyt *f* surface allows to sample the whole surface and to visualise all the possible locations of Pc.

Computational methods for studying protein complexes

Computational methods are essential for determining protein structures and studying the association of protein complexes, either in combination with experimental methods or in a purely theoretical manner. Furthermore, due to the limitations of experimental techniques, they represent a primary tool for the visualization of the encounter complex.[99,100]

In many structure determination studies, experimental restraints, such as those gained from X-ray crystallography, SAXS or NMR, are used as inputs for docking software packages, some of which are based on molecular dynamics procedures. The structures of plant,[46,62] *Ph. laminosum*[54] *Nostoc*[50] and *Pr. hollandica*[32] Pc-Cyt *f* complexes were obtained by rigid-body docking of the proteins, on the basis of experimentally obtained paramagnetic restraints and chemical shift perturbations. The experimental restraints were the driving force for the association of proteins, rigid-body dynamics was applied at each docking step and the structures were saved according to the energy values calculated for each complex. The same concept is applied in the ensemble docking procedure. In this case an ensemble of a variable number of elements is concurrently docked to the interaction partner, which is considered as single and fixed object. This method has been successfully employed to study intermolecular motions[101,102] and for the visualization of various kinds of encounter complex, such as DNA-protein[103,104] and protein-protein complexes.[27,92,105] This method is discussed further in Chapter III.

Purely theoretical approaches have supported both structural and functional experimental studies and in many cases they have enriched the understanding of ET pathways[56-58] and ET rates[61,68,71,106-108] in the Pc-Cyt *f* complex. Brownian dynamics (BD) and Monte Carlo (MC) simulations represent the most widespread methods used for studying protein-protein interactions. A mobile diffusing protein is docked to a target molecule under the influence of an electrostatic field and Brownian motions[109] or Monte Carlo sampling,[57] respectively. Both techniques allow the screening of a large number of conditions in a relatively short time. In this way, different mutations or ionic strength effects can be included in the simulations providing an extensive data set, which is hard to achieve experimentally. Assuming that the formation of the encounter complex is mainly driven by electrostatics, those methods allow to study the different configurations of the encounter intermediates. The combination of experimental NMR data and Monte Carlo simulations was successfully employed in the visualization of the encounter complex formed by another pair of ET proteins, cytochrome *c* and cytochrome *c* peroxidase, for which also the crystal structure had been obtained.[110] This encounter complex has proven to be absolutely dependent on electrostatic forces, in agreement with the theoretical predictions.[28] Furthermore, mutagenesis of the interface residues involved in the interaction, caused a change in the equilibrium between encounter and final state populations.[31]

Thesis outline

This thesis focuses on the characterization of the dynamic aspects of the transient complex of Cyt *f* and Pc in order to elucidate the nature of the mechanisms behind protein complexes formation.
In chapter II the Cyt *f*-Pc complex from *Nostoc* was characterized by PRE NMR spectroscopy, revealing a high degree of dynamics within the complex. The influence of spin label in the binding was also determined by chemical shift perturbations analysis.
Chapter III describes the visualization of the encounter state of the *Nostoc* Cyt *f*-Pc complex achieved through the combination of PRE NMR spectroscopy, Ensemble docking and Monte Carlo simulations. The role of hydrophobic interactions in the encounter complex formation has been experimentally demonstrated and a new model for protein complexes formation has been proposed.
The effect of electrostatic interactions on dynamics of the complex has been studied in the cross complex formed by *Nostoc* Cyt *f* and *Phormidium laminosum* Pc and it is presented in Chapter IV.
Finally, the results of this thesis have been contextualized in terms of encounter complex and its role in the association of ET systems (Chapter V).

Chapter II

The complex of cytochrome *f* and plastocyanin from *Nostoc sp*. PCC 7119 is highly dynamic

Adapted with permission from
Scanu S., Foerster J.M, Finiguerra M.G., Shabestari M.H., Huber M. and Ubbink M. (2012) The complex of cytochrome *f* and plastocyanin from *Nostoc sp.* PCC 7119 is highly dynamic. *Chembiochem*. **13**, 1312-1318. Copyright 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Abstract

Cyt *f* and Pc form a highly transient complex as part of the photosynthetic redox chain. The complex from *Nostoc sp.* PCC 7119 was studied with NMR relaxation spectroscopy with the aim to determine the orientation of Pc relative to Cyt *f*. Chemical shift perturbations analysis showed that the presence of spin labels on the surface of Cyt *f* does not significantly affect the binding of Pc. The paramagnetic relaxation enhancement results are not in agreement with a single orientation of Pc indicating that multiple orientations must occur and suggesting that the encounter state represents a large fraction of the complex.

Introduction

The concept of protein-protein complex formation is evolving towards a view in which an encounter state is in dynamic equilibrium with the well-defined, specific complex.
The initial approach of the proteins and subsequent formation of the encounter state are thought to be mainly driven by long-range electrostatic forces, whereas the well-defined complex is stabilized by short-range interactions, like hydrogen bonds and van der Waals forces.[21] Until recently it was not possible to characterize the encounter state experimentally. However, several existing methods have been adapted for this purpose, like double-mutant cycles combined with measurements of association kinetics,[111] flash photolysis kinetics[112] and PRE NMR spectroscopy.[84,113] The first complex of ET proteins characterized by this approach was that of cytochrome *c* (Cyt *c*) and cytochrome *c* peroxidase (CcP). The solution structure of this complex has been determined by PRE NMR[114] showing that Cyt *c* has the same orientation within the complex as in the crystal structure.[110] At the same time the PRE data provided evidence for dynamics within the complex, suggesting that the encounter complex was significantly populated. In combination with Monte Carlo docking the encounter state could be visualized and its fraction established to be 30%.[28] This approach opens the door for the characterization of the encounter state in other transient redox complexes.[20]
Pc and Cyt *f* form a redox complex in oxygenic photosynthesis. Pc shuttles electrons from Cyt *f* of the b_6f complex to P700 in PSI. The surface charge properties of Pc and Cyt *f*, which vary significantly between the different species, influence the relative orientation of the interaction partners in the well-defined complex. Two general orientations have been described, dubbed "side-on" and "head-on". The "side-on" orientation has been observed in plant complexes.[115,116] The plant proteins exhibit a favourable electrostatic interaction due to the presence of negatively and positively charged patches of amino acids on Pc and Cyt *f*, respectively. The patches align the long sides of Pc and Cyt *f*, enabling rapid ET by bringing a hydrophobic patch on Pc close to the haem in Cyt *f*.[115] In the cyanobacterial complex from *Phormidium laminosum*, Pc approaches Cyt *f* "head-on".[54] Within the complex, Pc is oriented perpendicular to the haem plane and only its hydrophobic patch participates in the interaction. Electrostatics play a smaller role in *Phormidium* than in plants,[117] although kinetics studies[118,119] suggested charge interactions contribute to the formation of the encounter state. In the cyanobacterial complexes from *Nostoc sp.* PCC 7119[120] and *Prochlorothrix hollandica,*[32] where the charge distribution is reversed compared to that in plants, again the "side-on" orientation was observed. The solution models of the complexes have been determined by rigid-body docking of the structures of the individual proteins on the basis of binding CSPs and intermolecular PCSs of Pc nuclei induced by the paramagnetic, oxidized iron of Cyt *f*.[115] In the case of the *Nostoc* complex site-directed mutagenesis studies investigating the influence of charges on the kinetics of the complex formation, highlighted how the loss of either positive charges on Pc[121] or negative charges on Cyt *f*[122] causes a decreased association rate constant. It could be shown that for Pc several

charges are pivotal for the interaction.[121,122] On the other hand, the charges on Cyt *f* are more spread over the surface and no "hot spots" were identified either in *Nostoc*[122] or *Phormidium,*[119] suggesting that the encounter complex may have an important role in these complexes. To obtain independent restraints for the refinement of the well-defined state and to establish whether the encounter state is significantly populated, the Pc-Cyt *f* complex from *Nostoc* was studied by PRE NMR spectroscopy. The data cannot be described by the structure determined by pseudocontact shifts alone, or indeed by any single structure, indicating that the encounter ensemble must represent quite a significant fraction of the complex.

Experimental section

Protein production and purification

The plasmid pEAP-WT containing the gene encoding *Nostoc sp.* PCC 7119 Pc was kindly provided by Prof. Dr. Miguel A. De la Rosa (University of Seville). The leader sequence, consisting of 34 amino acids, was removed in order to achieve cytoplasmic expression of the mature Pc (as defined in Uni*P*rot entry O52830). An N-terminal Met residue was added to initiate translation. This construct was obtained by subcloning with PCR, using the following primers according to the Cloning Standard Protocol (Qiagen®).

FWD: 5'- ctgtgcaa**ccatgg**aaacatacacagtaaaactaggtagcg -3'
REV: 5' -ctgtgcaa**ctcgag**ttagccggcgacagtgattttacc – 3'.

The restriction site for *Nco*I and *Xho*I were introduced in the forward and reverse primers, respectively, and are indicated with bold letters. The former comprises the ATG codon for initiation Met residue. The amplified gene and the vector pET28a were doubly digested with these enzymes before the ligation step. The construct (pSS01) was verified by DNA sequencing.
Uniformly ^{15}N-labeled Pc was produced in *E. coli* BL21 freshly transformed with pSS01. A single colony was inoculated in lysogeny broth (LB, 2 mL) with kanamycin (25 mg/L) and cultured until the OD_{600} reached 0.6. Fifty microlitres were used to inoculate ^{15}N M9 minimal medium (50 mL) with $^{15}NH_4Cl$ (0.3 g/L) as only source of nitrogen and kanamycin (25 mg/L), which was incubated overnight. Five millilitres of culture were transferred into ^{15}N minimal medium (0.5 L), and incubated until the OD_{600} reached 0.6. All cultures were incubated at 37°C, 250 rpm. Then, the expression of the gene encoding Pc was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM) and the temperature was decreased to 22°C. The cells were harvested after 20 hours by centrifugation at 6400 g for 20 min at 4°C. The pellet was resuspended in sodium phosphate (NaP_i, 10 mL, 1 mM, pH 7). Phenylmethylsulfonyl fluoride (PMSF, 1 mM), DNase (0.2 mg/ mL) and $ZnCl_2$ (250 μM) were added. Cells were lysed using a French Press. The cell lysate was cleared by ultracentrifugation at 210000 g for 30 min at 4°C and the supernatant was dialyzed overnight against NaP_i (1 mM, pH 7), $ZnCl_2$ (25 μM). The solution was

cleared by ultracentrifugation and loaded onto a carboxymethyl (CM) cellulose Sephadex C-50 column (GE Healthcare) equilibrated with the same buffer, in the absence of zinc. The elution was carried out with a gradient of NaP_i (1-25 mM, pH 7). The fractions containing Pc were loaded once again on the same column and eluted under the same conditions. The concentration of the protein was determined by absorbance spectroscopy using ε_{280} = 5 $mM^{-1}cm^{-1}$. The yield of pure protein was (10 mg/L) of culture. The absence of Cu was verified by UV-vis spectroscopy by the absence of the characteristic band at 595 nm under oxidizing conditions. The presence of Zn was verified by atomic absorption spectroscopy.

The pEAF-WT plasmid, containing the gene of the soluble domain (residue 1-254) of *Nostoc sp.* PCC7119 Cyf *f* was kindly provided by Prof. Dr. Miguel A. De la Rosa (University of Seville). The pEAF-WT plasmid was used as template to obtain Cyt *f* mutants. The mutations to cysteine were introduced by using the QuikChange Site-Directed Mutagenesis kit (Stratagene). The primers used for the mutations at the positions N71 and Q104 were described before.[123] The primers employed for the introduction of a cysteine at the position S192 were:

FWD: 5' –ggcgaagatggttgcgttaaatatttagt**c**gacatc-3'
REV: 5' – gatgtc**g**actaaatatttaacgcaaccatcttcgcc-3'

For S192C a silent mutation (bold) was designed to introduce an extra Sal I restriction site, located close to the *3'* end of the forward primer. The codon-changing mutations are shown underlined. The mutant genes were verified by DNA sequencing.

Truncated Cyt *f* was produced in *E. coli* MV1190 (Δ(*lac-pro*AB), *thi*, *sup*E, Δ(*srl-rec*A) 306::Tn10 (tet^r) [F′:*tra*D36, *pro*AB+, $lacI^qZ\Delta M15$]), transformed with pEAF-WT or mutant plasmids and co-transformed with pEC86,[124] which contains a cassette for *c*-type cytochrome overexpression. A single colony of co-transformed *E.Coli* MV1190 was inoculated into LB (50 mL) containing chloramphenicol (20 mg/L) and ampicillin (100 mg/L), and cultured at 37°C, 250 rpm for 5-6 h. Five millilitres of culture were used to inoculate LB (1.7 L) with the same antibiotics in a 2 L Erlenmeyer flask. The culture was incubated at 25°C, 150 rpm for 20 h before further addition of chloramphenicol (20 mg/L) and ampicillin (100 mg/L). After again two hours incubation, gene expression was induced with IPTG (1 mM). The cells were harvested 96 hours after induction. The purification of the protein was performed as previously reported.[121] Dithiothreitol (DTT, 3 mM) was added to the buffers during the purification to prevent the dimerisation of the cysteine mutants. It was removed immediately before spin labelling by buffer exchange on a PD10 column (GE Healthcare), equilibrated with 2-(*N*-morpholino) ethanesulfonic acid (MES) buffer (20 mM, pH 6). The ferrous form of the Cyf *f* cysteine mutants was used for the attachment of (1-Acetoxy-2,2,5,5-tetramethyl-δ-3-pyrroline-3-methyl) methanethiosulfonate (MTS) or (1-Oxyl-2,2,5,5-tetramethyl-δ-3-pyrroline-3-methyl) methanethiosulfonate (MTSL), with a protein concentration ranging from 20 to 80 μM. A 20-fold molar excess of MTS(L) was added and the solution was incubated for one hour on ice. A 100-fold molar excess of $K_3[Fe(CN)_6]$ was then added to oxidize the haem iron and prevent reduction of the nitroxyl group or the

disulphide bridge by the ferrous haem.
The sample was concentrated by ultrafiltration to a volume of 0.5 mL and loaded on a Superdex75 gel filtration column (GE Healthcare) equilibrated with MES buffer (20 mM, pH 6). The fractions containing MTS(L)-Cyt *f* were concentrated and the buffer was exchanged by ultrafiltration to MES (20 mM, pH 6), $K_3[Fe(CN)_6]$ (0.5 mM). The attachment of the spin label was verified by mass spectrometry and the presence of the nitroxyl radical was checked by EPR spectroscopy. The concentration of the protein was determined by absorbance spectroscopy using $\varepsilon_{556} = 31.5\ mM^{-1}cm^{-1}$ for ferrous Cyt *f*.

NMR experiments

All NMR samples contained MES (20 mM, pH 6) and 6% D_2O for lock. Cyt *f* was kept in the ferric state by addition of $K_3[Fe(CN)_6]$ (50 μM). The pH of the sample was adjusted with small aliquots of HCl (0.5 M) and NaOH (0.5 M). For the chemical shifts perturbation experiments Cyt *f* was titrated into Zn-substituted ^{15}N Pc (50 μM). Spectra were recorded at multiple Cyt *f*:Pc molar ratios (0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 2.5, 5.0). Samples for PRE measurements contained Cyt *f* (66 μM) labelled with either MTS or MTSL and Zn-substituted ^{15}N Pc (200 μM). All NMR spectra were recorded at 298 K on a Bruker Avance III 600 MHz spectrometer equipped with a TCI-Z-GRAD CryoProbe. The ^{1}H-^{15}N HSQC spectra were acquired with 1024 and 80 complex points in the direct and indirect dimensions, respectively.

Data analysis

The NMR spectra were processed with NmrPipe[125] and analyzed with CcpNMR.[126] The assignments for the Zn ^{15}N Pc amide resonances were kindly provided by Dr. Mathias A. S. Hass. The assignments for the residues K6 and V29 could not be made because of overlap of the corresponding peaks in the HSQC spectra. The chemical shift perturbations ($\Delta\delta_{bind}$) of Pc resonances due to complex formation with Cyt *f* were plotted against the molar ratio Cyt *f*/Pc (*R*). Note that the perturbations include both the effect of binding and the PCS caused by the ferric haem iron in all samples. The entire perturbation was used in the analysis and the PCS was not used separately in this study.
The corresponding titration curves were fitted in OriginLab 8.1 (www.originlab.com) with a non-linear least square fit to a 1:1 binding model, Equation 2.1.[127]

$$\Delta\delta_{bind} = \frac{1}{2}\Delta\delta^{\infty}_{bind}\left[A - \sqrt{A^2 - 4R}\right] \quad A = 1 + R + \frac{P_0R + C_0}{P_0C_0K_a} \qquad (2.1)$$

Where $\Delta\delta^{\infty}_{bind}$ is the chemical shift perturbation for 100% bound Pc, P_0 is the starting concentration of Pc and C_0 is the stock concentration of Cyt *f*. A global fit with a single binding constant ($K_a = K_D^{-1}$) for the data of several residues was used. The binding maps were obtained by extrapolation of the $\Delta\delta_{bind}$ values at the 5:1 Cytf:Pc molar ratio of all residues to 100% bound Pc using the K_D. These

extrapolated perturbations were averaged for the nitrogen ($\Delta\delta_N$) and hydrogen ($\Delta\delta_H$) atoms of each amide, yielding $\Delta\delta_{avg}$, according to Equation 2.2:

$$\Delta\delta_{avg} = \sqrt{\frac{(\Delta\delta_N/5)^2 + \Delta\delta_H{}^2}{2}} \qquad (2.2)$$

The PREs were determined according to the procedure of Battiste and Wagner.[128] The intensity ratio I_p/I_d of the Pc resonances in the presence of MTSL-Cyt *f* (I_p) and MTS-Cyt *f* (I_d) were normalized by dividing them by the average value of the ten largest I_p/I_d values (1.13 for N71C and Q104C; 1.06 for S192C). The PRE (Γ_2) values were calculated according to the formula:

$$\frac{I_p}{I_d} = \frac{R_{2d}\exp(-\Gamma_2 t)}{R_{2d} + \Gamma_2} \qquad (2.3)$$

The transverse relaxation rates in the diamagnetic sample (R_{2d}) were calculated from the linewidth at half height obtained from a Lorentzian peak fit in the direct dimension, by using MestReC (www.metsrelab.com). The symbol *t* denotes the time for transverse relaxation during the pulse sequence (9 ms).

Structure calculations

The PREs were converted into distances for structure calculations using Equation 2.4:

$$r = \sqrt[6]{\frac{\gamma^2 g^2 \beta^2}{\Gamma_2}\left(4\tau_c + \frac{3\tau_c}{1+\omega_h^2\tau_c^2}\right)} \qquad (2.4)$$

Where *r* is the distance between the oxygen atom of MTSL and the Pc amide proton, γ is the proton gyromagnetic ratio, *g* is the electronic g-factor, β is the Bohr magneton, ω_h is the Larmor frequency of the proton and τ_c is the rotational correlation time of the MTSL oxygen-proton vector. τ_c was taken to be 30 ns on the basis of the HYDRONMR[129] prediction of the rotational correlation time for the Pc-Cyt *f* complex. In the docking procedure this value gave rise to the lowest energy structures in comparison with τ_c values of 20 ns, 25 ns, 35 ns and 40 ns.
Three classes of restraints were included in the calculations: 1) For residues with I_p/I_d ratios < 0.1, including those for which the resonances disappeared from the spectrum, the upper bound distance limit was set to 14 Å; 2) For residues with I_p/I_d ratios > 0.95 the lower bound distance limit was set to 22 Å; 3) For residues with I_p/I_d ratios between 0.1 and 0.95 the distances calculated with equation 2.4 were used with upper and lower bounds of 4 Å. The structure calculations were done in Xplor-NIH.[130] Cyt *f* and Pc were both considered as rigid bodies, the coordinates of Cyt *f* were fixed, and Pc was allowed to move in a restrained rigid-body molecular dynamics calculation. The structure of the soluble domain of Cyt *f* used

for the calculation was taken from the crystal structure of the b_6f complex from *Nostoc* sp. PCC 7120, PDB entry 2ZT9.[131] The amino acidic sequences of Cyt *f* from *Nostoc* sp. PCC 7120 and sp. PCC 7119 are identical. Mutations and spin labels were modelled on the structure of Cyt *f*. Four conformations were used to represent the mobility of the spin label,[132] and the distances to the Pc nuclei were r^{-6} averaged for these MTSL conformers. The structure of Pc was taken from PDB entry 2CJ3. At each cycle Pc was placed at a random position and the protein was docked as rigid body only on the basis of the experimental restraints and a van der Waals repel function to avoid steric collision. Two hundred approaches were performed, yielding 155 structures with restraint energies below a given threshold. The ten lowest energy structures were selected and they showed an average r.m.s.d. difference of 0.83 Å to the mean structure.

Results and discussion

Characterisation of MTS-tagged Cyt *f*

To study the complex of Cyt *f* and Pc with PRE NMR spectroscopy, three sites for probe attachment were selected. The positions of the mutations were designed on the basis of the solution structure of the wild type complex determined by NMR spectroscopy on the basis of PCS and chemical shift perturbations.[120] The rationale of the work followed that of Volkov et al.[114] for the complex CcP and Cyt *c*, i.e. to obtain constraints for structure determination and improve the precision of the solution structure that was based on PCS. The residues that were mutated to cysteine were N71, Q104, and S192, which are located around the Pc binding site (Figure 2.1). In order to preserve the overall electrostatic potential in the complex only polar, neutral amino acid residues were selected.

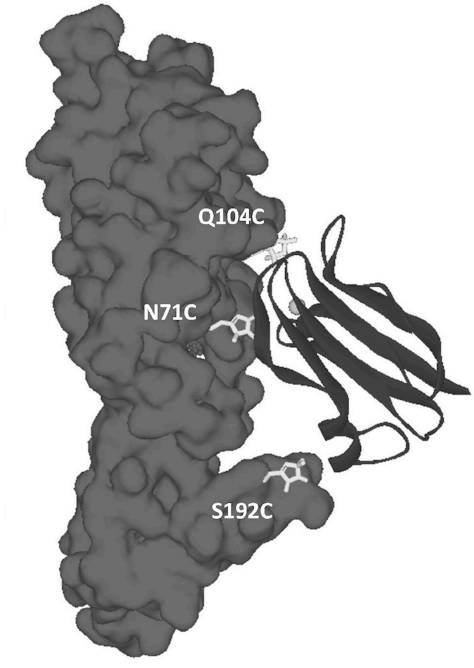

Figure 2.1. Location of the spin labels used in this study on the *Nostoc sp.* PCC 7119 Pc-Cyt *f* complex (PDB entry 1TU2, model 1[133]). Pc is shown in ribbons with the copper as a sphere. Cyt *f* is shown as surface. The spin labels were modeled on the structure and are shown as sticks.

Pc was produced in a cytoplasmic expression system (see the Experimental Section) in the ^{15}N enriched form with a Zn^{II} ion in the copper binding site to eliminate the paramagnetic effect of Cu^{II} and possible interference from electron transfer reactions.[134] To establish whether the introduction of a probe interferes with the Pc-Cyt *f* interaction, chemical shift analyses were carried out for all variants. First, Zn-substituted ^{15}N Pc was titrated with wt Cyt *f* and HSQC spectra were acquired

at each titration point. The binding constant was obtained by fitting the chemical shift perturbation curves for the most affected amide groups (Figure 2.2A), yielding K_D 8(3) x 10^{-5} M, similar to the reported values of 3.8 (0.1) x 10^{-5} M for Zn-Pc[135] and 6.2(0.9) x 10^{-5} M for Cd-substituted Pc.[135]

Also the binding map is similar, with the largest perturbations observed for the residues L14, G94 and A95, corresponding to the hydrophobic interaction patch, and H92 and R93, belonging to the basic patch (Figure 2.2B).

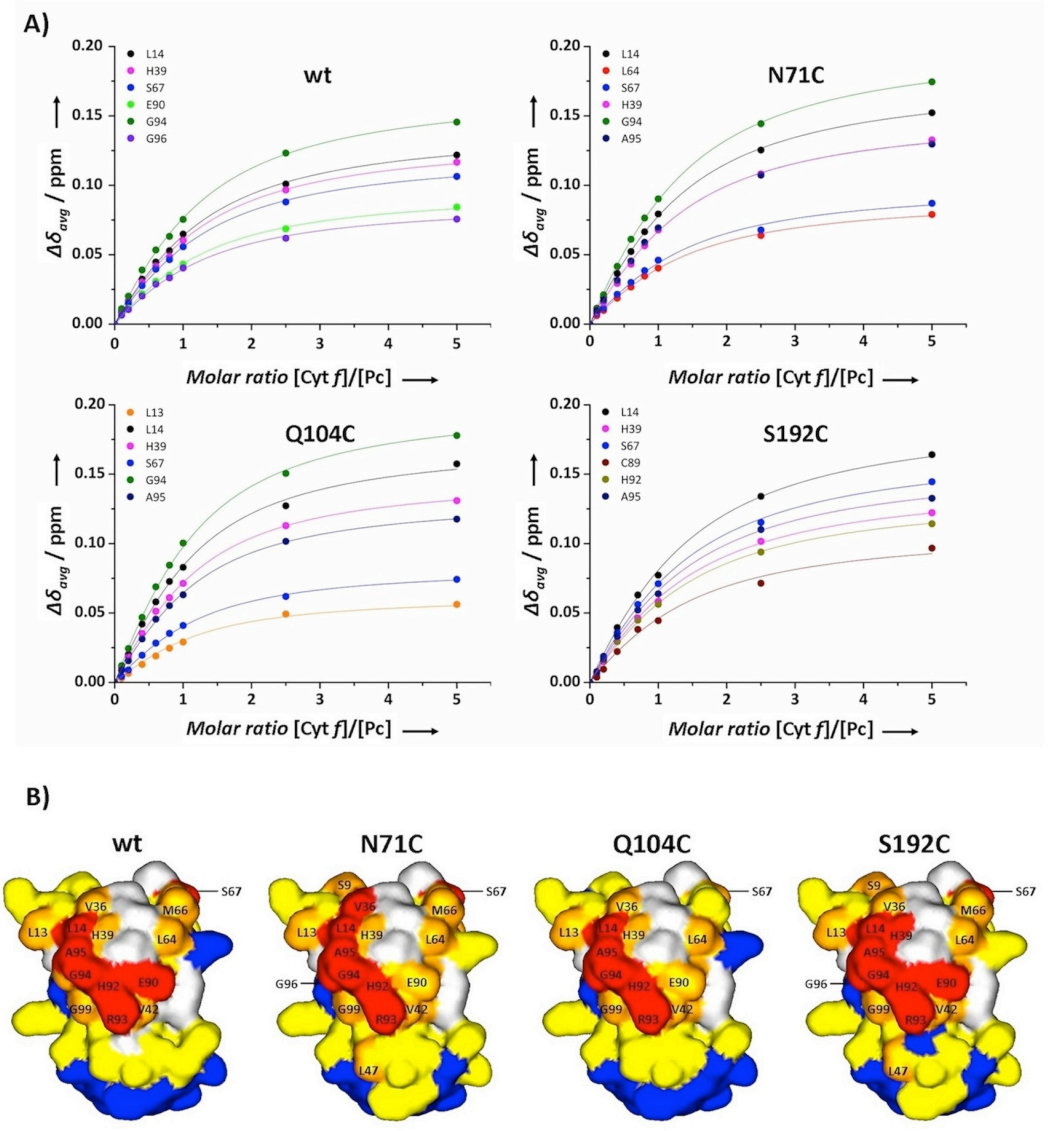

Figure 2.2. The interaction of *Nostoc* Zn-substituted Pc with wt Cyt *f* and MTS conjugated variants. A) Binding curves for selected residues were fitted globally to a 1:1 binding model (equation 2.1). B) Chemical shift perturbation maps of Zn-substituted Pc in the presence of wild type and MTS-conjugated Cyt *f*, colour-coded on a surface model of Pc (PDB-entry 2CJ3), with red, $\Delta\delta_{avg} \geq 0.10$ ppm; orange, $\Delta\delta_{avg} \geq 0.05$ ppm; yellow, $\Delta\delta_{avg} \geq 0.02$ ppm; blue, $\Delta\delta_{avg} < 0.02$ ppm. Prolines and residues with overlapping resonances are in grey.

Similar titrations of Pc and the cysteine mutants of Cyt *f* conjugated with the diamagnetic control label MTS yielded the dissociation constants listed in Table 2.1. The binding curves and maps are shown in Figure 2.2. Clearly, the mutation and attachment of MTS have very little effect on the affinity and the binding map.

Table 2.1. Dissociation constants of the complexes formed by *Nostoc* (Zn) Pc with wt and MTS-conjugated Cyt *f*. The errors are indicated in parentheses.

Cyt *f* mutant	K_D * 10^{-5} (M)
wild type	8 (3)
N71C-MTS	4 (1)
Q104C-MTS	3 (1)
S192C-MTS	4 (1)

Paramagnetic relaxation enhancements of Pc nuclei

The aim of this study was to gather distance restraints from PREs to refine the published solution structure. The residues selected for mutation to cysteine and tagging with the spin label are located around binding site for Pc on Cyt *f* in the solution structure model. The spin labels at these positions are thus expected to yield PRE of nuclei on different sides of Pc. For this purpose, the spin label MTSL was attached to the three Cyt *f* mutants and the tagged protein was added to Pc to a Cyt *f*/Pc molar ratio of 0.3. Under these conditions, the fraction of bound Pc is 24%. Large PREs were observed already at this ratio for numerous Pc amide groups with each variant of Cyt *f*-MTSL, illustrated in Figure 2.3.

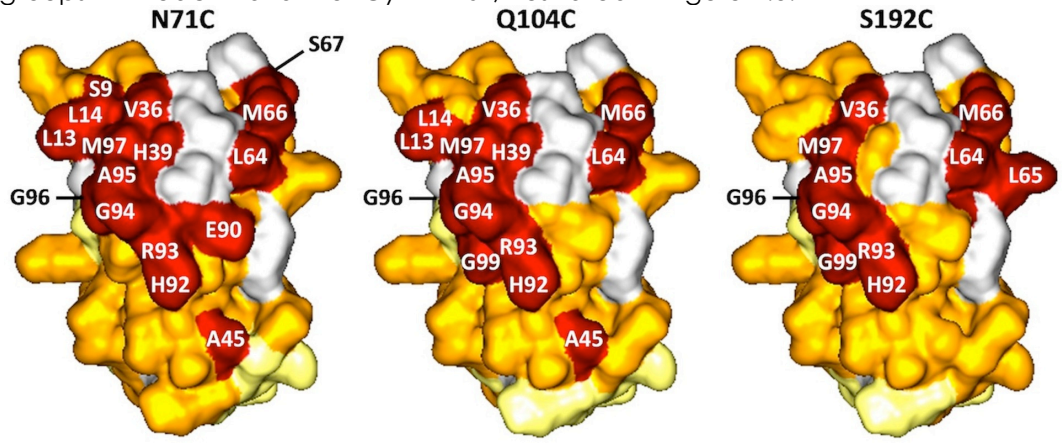

Figure 2.3. PRE maps of Zn-substituted Pc bound to MTSL-conjugated Cyt *f*, colour-coded on a surface model of Pc (PDB-entry 2CJ3) according to the three classes of restraints defined for the docking: red corresponds to residues with $I_p/I_d<0.10$; orange to $0.10<I_p/I_d<0.95$ and yellow to $I_p/I_d>0.95$. Prolines and residues with overlapping resonances are in grey.

Some resonances were broadened beyond detection. It is clear from the Cyt *f*

titrations that Pc binding is in the fast exchange regime, so an observed PRE is a weighted average of free Pc (no PRE), the encounter state (the AB* ensemble) and the final complex, AB, (Equation 2.5).

$$PRE_{obs} = f_1^{\text{free Pc}} \times 0 + f_2^{AB^*} \langle PRE \rangle^{AB^*} + f_3^{AB}\, PRE^{AB} \qquad f_1 + f_2 + f_3 = 1 \qquad (2.5)$$

The fraction of the free Pc (f_1) is 0.76 and that of bound Pc (f_2+f_3) is 0.24. By dividing the observed PRE by 0.24, the PRE for 100% bound Pc is obtained. These extrapolated PREs are plotted in Figure 2.4 (green symbols) against the Pc residue number.

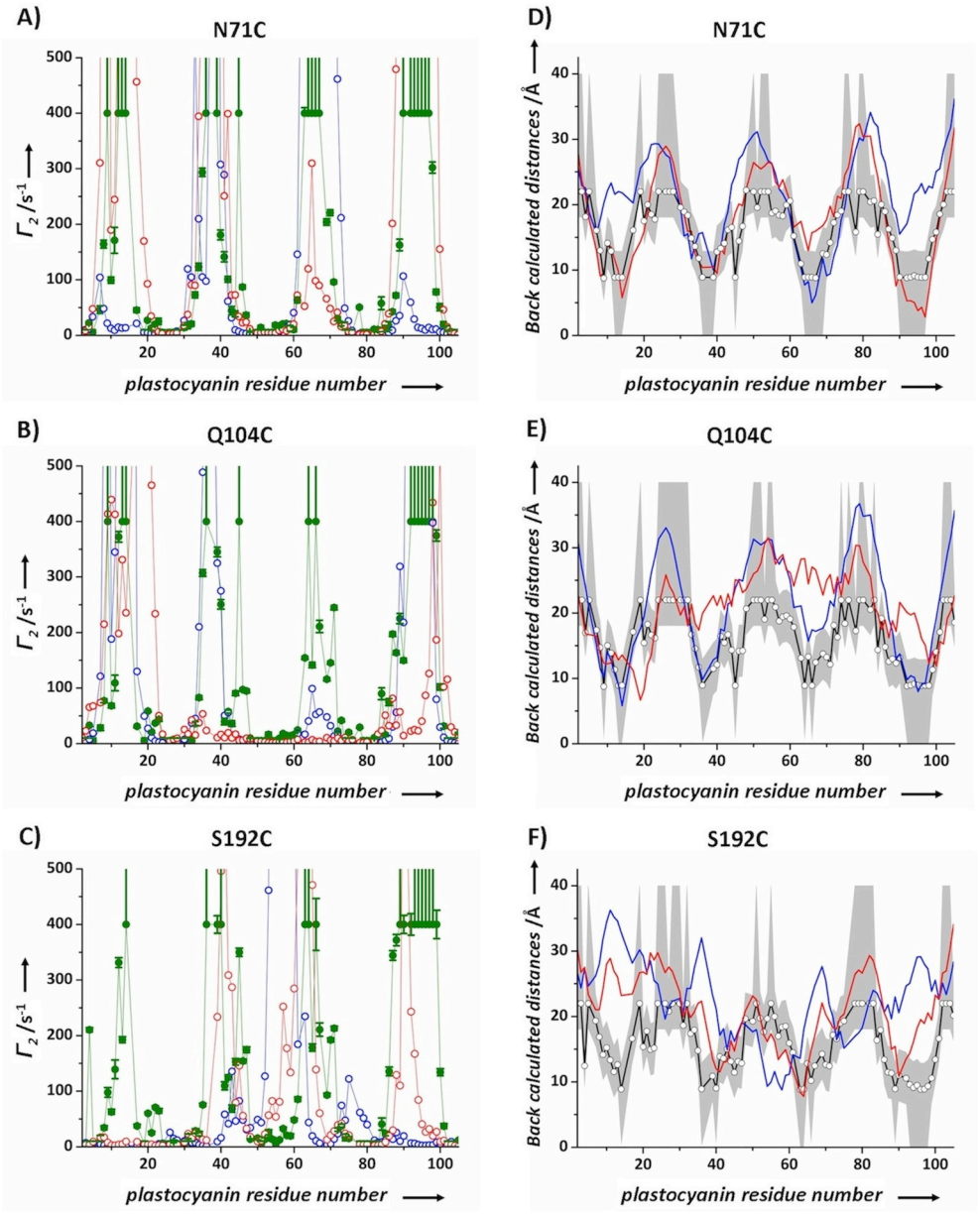

Figure 2.4. Left panel (A-C), observed and predicted PRE values for amide protons in Pc bound to MTSL conjugated Cyt *f* variants N71C (A), Q104C (B) and S192C (C). The observed PREs were extrapolated to represent the 100% bound state of Pc (green dots). PREs at 400 s^{-1} represent lower limits. The PREs calculated from the NMR solution structure based on PCS (PDB entry 1TU2, model 1[133]) are shown as blue symbols. The PREs calculated from the NMR solution structure based on PREs are shown as red symbols. Right panel (D-F), experimental and back calculated distances between Pc amide protons and MTSL conjugated Cyt *f* variants N71C (D), Q104C (E) and S192C (F). The white circles and black line represent the distances calculated from the experimental PREs (which were extrapolated to 100% bound Pc), where the grey area indicates the error margins. The distances derived from the NMR solution structure based on PREs are shown as a red line. The distances derived from the NMR solution structure based on PCS (PDB entry 1TU2, model 1[133]) are shown as a blue line.

Strikingly, the patterns are qualitatively similar for the three spin label positions, indicating that the same patches of Pc are strongly affected. When the fraction of AB* is neglected ($f_2 \approx 0$), the PRE can be predicted from the model of the final complex. Using model 1 of PDB entry 1TU2, the PRE^{AB} values were predicted for each amide in Pc (Figure 2.4, blue symbols). Clearly, the model alone cannot account for the observed PREs. Also, docking calculations were performed using distances derived from the PREs as restraints. Apart from a van der Waals repel function to avoid steric collisions no other interactions were included. The ensemble of the ten best structures is shown in Figure 2.5 and compared with the model based on PCS. In both cases Pc is bound in the region close to the haem, but the orientation differs. However, also the PRE-based model on its own cannot account for the observed PREs and the back calculated distances (Figure 2.4, red symbols in the left panel and red line in the right panel, respectively).

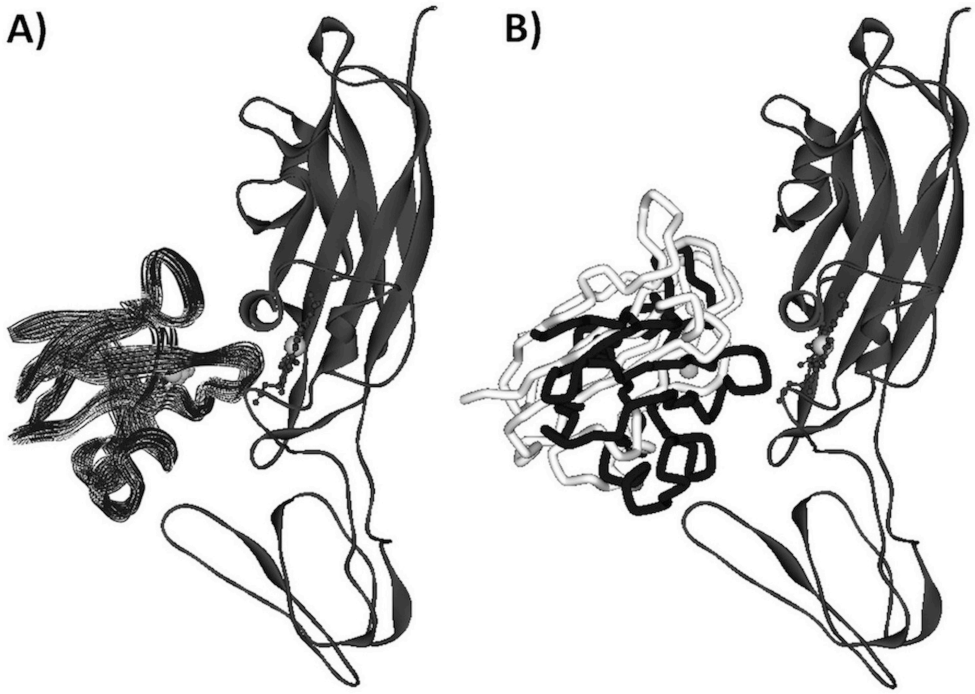

Figure 2.5. A) Model of the Pc-Cyt *f* complex obtained with PREs restraints. Cyt *f* is shown in ribbons and Pc as Cα trace. The ten lowest energy structures are visualized. B) Overlay of the orientations of Pc in the NMR solution structure based on PREs (shown as black Cα trace) and the NMR solution structure based on PCS (PDB entry 1TU2, model 1 [133], light grey).

Thus, it can be concluded that a single orientation is insufficient to describe the Pc-Cyt *f* complex. It is now well-established that PREs are very sensitive to lowly populated states. The poor fit between the PRE data and the modelled structure indicates that other orientations of Pc within the complex contribute to the PRE data. This conclusion is also borne out by the similarity of the PRE maps in Figure 2.3 for the Cyt *f* variants. If the Pc were in a single orientation in the complex, different patches of Pc residues would have been affected by PRE, because the spin labels are located around the binding site (see, for example, the red symbols in Figure 2.4). Yet, for each spin label position the same Pc surface region was affected by PREs, this also matches the side with the largest chemical shift perturbations. These observations suggest that Pc samples a large area of the surface of Cyt *f* with its hydrophobic patch. Our results are in accord with kinetic experiments,[121,122] which indicated that the interaction site of Pc depends on a few specific residues, whereas for Cyt *f* the residues relevant in the association are more spread over the protein surface. The formation of the encounter complex reduces the dimensionality of the diffusional search for the binding site that enables rapid ET.[21] It has been suggested that the population balance between the encounter state and the well-defined state depends on whether rapid ET can occur in encounter state orientations.[28] In complexes of small proteins the redox centres can get sufficiently close for ET in many of the protein orientations, but in larger complexes fast ET can only occur through certain areas of the protein surface, requiring the formation of a well-defined complex. Both Cyt f and Pc have an elongated shape and the metal is close to only a small part of the surface. Thus, it seems likely that a degree of specificity in the interaction is required in this complex to be active. Earlier studies suggested that the degree of dynamics varies between Cyt *f*/Pc complexes. Those of *P. laminosum*[54] and *Pr. hollandica*[32] appeared to be particularly dynamic. From the data presented in this study the fraction of the encounter complex cannot be established, but it is clear that it is significant, even though the complex from *Nostoc* would be categorized as well-defined given the earlier NMR data, the intermolecular PCS from the Cyt *f* haem to Pc.[120] Both PCS and PRE are sensitive to minor states populating the encounter complex in the case that the minor state experiences a much stronger paramagnetic effect than the major state. In the opposite case, when the major orientation is most affected by the paramagnetism, the presence of minor states may well go unnoticed, because it only leads to a small reduction of the observed effect. Here, the PCSs are large in particular for the major state, close to the paramagnetic haem, whereas the PREs will be dominated by those orientations that bring the nucleus close to the spin label, and thus, the PREs describe better the combination of the final complex and the encounter ensemble. Therefore, a good fit of the PCS data could be obtained with a single structure in the study of Diaz et al.,[120] while the same structure is insufficient to account for all PREs. More extensive spin labelling covering a large area of the surface will enable a detailed description of the encounter complex, as was shown for the ET complex of Cyt *c* and CcP.[28] Such experiments are presented in Chapter III.

Chapter III

Role of hydrophobic interactions in the encounter complex formation of plastocyanin and cytochrome *f* complex revealed by paramagnetic NMR spectroscopy

Adapted with permission from:
Scanu S., Foerster J.M., Ullmann G.M. and Ubbink M. (2013) Role of hydrophobic interactions in the encounter complex formation of plastocyanin and cytochrome *f* complex revealed by paramagnetic NMR spectroscopy. *J Am Chem Soc* **135**, 7681-7692. Copiright 2013 American Chemical Society.

Abstract

Protein complex formation is thought to be at least a two-step process, in which the active complex is preceded by the formation of an encounter complex. The interactions in the encounter complex are usually dominated by electrostatic forces, whereas the active complex is also stabilized by non-covalent short range forces. Here, the complex of Cyt *f* and Pc, electron transfer proteins involved in photosynthesis, was studied using paramagnetic relaxation NMR spectroscopy. The labeling of Cyt *f* with nine spin labels shows that a large part of the protein surface area is sampled by plastocyanin. In contrast, plastocyanin is always oriented with its hydrophobic patch toward Cyt *f*. The complex was visualized using ensemble docking, showing that the encounter complex is stabilized by hydrophobic as well as electrostatic interactions. The results suggest a model of electrostatic pre-orientation before the proteins make contact, followed by the formation of an encounter complex that rapidly leads to electron transfer active conformations by gradual increase of the overlap of non-polar surface areas on Cyt *f* and plastocyanin. In this model the distinction between the encounter and active complexes vanishes, at least in the case of electron transfer complexes, which do not require a high degree of specificity.

Introduction

A general model for protein-protein interactions describes protein association as a stepwise process in which the formation of the final complex is preceded by that of a transient, lowly populated state encounter complex.[22] In the first step of association the proteins are steered towards the encounter complex by diffusion and long-range electrostatic interactions. In the encounter state proteins show few specific interactions. They rather tend to assume multiple orientations to sample the surface of the partner and reduce the dimensionality of search of the specific binding site.[21] The final complex is dominated by short-range, specific interactions, which stabilize it in a single orientation. An encounter complex will not always proceed towards the final complex. In some cases it is futile and will dissociate again.[24,136] The applicability of this model has been theoretically and experimentally demonstrated for complexes with electrostatic assisted association, both when the interaction partners showed high overall charge complementarity[23,137] and when opposite charges are more localized in specific regions on the surface of the reactants.[28,109] At the same time, given the wide variety in the electrostatic surface properties of proteins, this model cannot readily be generalized for all protein complexes. For complexes in which the interaction partners do not present charge complementarity or apparent dipolar interactions, a desolvation-mediated association has been suggested on the basis of theoretical work,[138,139] in which hydrophobic interactions guide both encounter complex formation and stabilization of the final complex. Experimental evidence of a predominantly hydrophobic-driven binding event is rare, although some has been reported.[54,140]

To investigate the contribution of the different forces involved in the molecular recognition process for transient complexes involved in electron transfer (ET), the complex formed by Pc and Cyt *f* from the cyanobacterium *Nostoc sp.* PCC 7119 was studied. Pc and Cyt *f* are redox partners in oxygenic photosynthesis in plants, green algae and cyanobacteria. Pc transfers electrons from Cyt *f* of the b_6f complex to PSI.[39] In both proteins the redox active sites are buried below extensive hydrophobic surface patches, which form the specific binding site. The overall electrostatic properties of the proteins vary significantly between different species and influence the final orientation of the complex. In plants, the final complex has been shown to be electrostatically stabilized in a "side-on" orientation by complementary localized charges on the protein surfaces, negative in Pc and positive in Cyt *f*.[44,62] Complementary charges tilt Pc toward the long side of Cyt *f* and align the hydrophobic binding sites, thus facilitating the ET reaction. In the cyanobacterium *Phormidium laminosum*, the final complex assumes a "head-on" orientation.[54] Pc is oriented perpendicular to the haem plane and comes into contact with Cyt *f* only with the hydrophobic patch. The "side-on" orientation was also observed in the complexes from the cyanobacteria *Nostoc*[98] and *Prochlorothrix hollandica*,[32] in which the charges are inverse compared to the plant counterparts, being positive in Pc and negative in Cyt *f*. Site-directed mutagenesis of key interface residues important for the overall electrostatic potential of the proteins from *Nostoc*, demonstrated that electrostatic interactions heavily regulate the kinetics of complex formation.[8,66]

Interestingly, the loss of negatively charged residues in Cyt *f*, in which charges are spread over a large part of the surface instead of being localized in a specific region, caused only small changes of the association rate constant,[66] whereas mutations of positively charged residues in a conserved region in Pc showed these charges to be fundamental for fast association.[8] The surface charge properties of Pc and Cyt *f* appear to influence the degree of dynamics within the complexes.[107] The *Ph. laminosum* and *Pr. hollandica* complexes showed to be highly dynamic.[32,54] The solution structures of the above mentioned complexes have been obtained by taking advantage of the pseudocontact shift (PCS) caused by the paramagnetic haem iron of Cyt *f* on backbone amide protons of Pc. PCS from the haem are not very sensitive to dynamics although the presence of many different Pc orientations will cause a decrease in the observed average PCS. In contrast, a recent paramagnetic relaxation enhancement (PRE) NMR study on the *Nostoc* system revealed that the Pc-Cyt *f* complex is more dynamic than was suggested by PCS, indicating the presence of a significantly populated encounter state.[29]

PRE NMR spectroscopy has proven to be a sensitive technique for the detection and visualization of lowly populated intermediates in protein-DNA[96] and protein-protein complexes.[28,140,142] PREs arise from magnetic dipolar interactions between the unpaired electron of a paramagnetic center and the observed nucleus, which causes an increase in the relaxation of the latter. Due to the large magnetic moment of the unpaired electron and the inverse sixth power distance dependence of the PRE, it is very large for nuclei that spend time in close proximity of the paramagnetic center. The sensitivity of PRE for lowly populated states is due to the fact that in the NMR fast exchange regime the observed PRE rate is a population weighted average of all species present in solution.[84] If in the lowly populated state the nucleus is close to the paramagnetic center, a PRE can be detected on the exchange averaged signal, even if the population is as low as 1%. Intermolecular PREs can thus be used to investigate transient intermediates in protein-protein complexes.[140] The observed PREs provide explicit qualitative evidences of the presence of the encounter state, but they do not provide a complete description of the encounter complex. Furthermore, the visualization of the encounter complex is an 'inverse' problem since many possible solutions can correspond to the observed PREs. To depict the encounter complex, experimental data need to be supported by theoretical models, generated by computational approaches. In the ensemble docking approach, multiple conformers of a protein are simultaneously docked to the other protein on the basis of the experimental PRE.[140] The encounter complex is visualized as an ensemble of orientations that fit the experimental restraints. In purely theoretical methods, such as Brownian dynamics (BD) and Monte Carlo (MC) simulations, in which proteins are docked only on the basis of electrostatic interactions,[57,109] the encounter complex is given as a distribution of the favorable electrostatic orientations.[28]

In this study, we report the visualization of Pc-Cyt *f* encounter complex and the characterization of this complex by the combination of PRE NMR, MC simulations and ensemble docking. The comparison of the results from the different approaches indicates that long-range electrostatic interactions pre-orient Pc with the hydrophobic patch towards Cyt *f* and that Pc maintains the same orientation

while sampling the surface of Cyt *f*. The ensemble docking visualization of the encounter complex showed that the encounter complex is stabilized by hydrophobic as well as electrostatic interactions. We propose a model for Pc-Cyt *f* complex formation in which long-range electrostatic interactions pre-orient the unbound proteins before they make contact. Pc diffusively binds to Cyt *f* forming an extended encounter complex stabilized by the overlap of the respective non-polar surface areas and the encounter complex rapidly evolves to ET active conformations. The ambiguous distinction between the encounter and the active complex in this system will be discussed in the context of the physiological b_6f complex.

Experimental section

Protein production and purification

^{15}N enriched, Zn-substituted Pc was produced and purified as described before.[29] The concentration of the protein was determined by absorbance spectroscopy using $\varepsilon_{280} = 5\ mM^{-1}cm^{-1}$. The yield of pure protein was 10 mg/L of culture.
The pEAF-WT plasmid, containing the gene of the soluble domain (residue 1-254) of *Nostoc* sp. PCC7119 Cyf *f* was kindly provided by Prof. Dr. Miguel A. De la Rosa (University of Seville). Cyt *f* mutants were obtained using pEAF-WT plasmid as template for mutagenesis. The mutations to cysteine were introduced by using the QuikChange Site-Directed Mutagenesis kit (Stratagene) as described in Chapter II. The primers used for the mutations at the positions N71, Q104 and S192 were described before.[29,143] The primers employed for the introduction of a cysteine at the positions Q7, Q38, A63, Q125, S181 and Q242 are reported in Table 3.1.

Table 3.1. Nucleotide sequence of the primers used in site-directed mutagenesis of Cyt *f*. Codon-changing mutations are shown in bold, italic and underlined; silent mutations are in bold.

Mutation	Primer sequence
Q7C	FWD: 5'-gcatatcctttctgggcgcag***tgc***acttacccag-3'
Q38C	FWD: 5'-gcccacagaagttgaagttcct***tgc***tccgtactacccgacaccg-3'
A63C	FWD: 5'-ccagcgt**c**caacaagttggt***tgc***gatggctctaagg-3'
Q125C	FWD: 5'-cccgg**g**gaacagtat***tgc***gaaatcgtcttccctgttctttctcccaacccc-3'
S181C	FWD: 5'-gcgctgctgctaccggtacaatt***tgc***aagattgctaaacaaga**g**ggcg-3'
Q242C	FWD: 5'-ccctaacgttggtggtttcggt***tgc***ct**c**gacgcagaaattgttctcc-3'

In each primer a silent mutation (bold) was designed to remove or to introduce an extra restriction site. In the cases of Q7C and Q38C mutations, the codon-changing mutations (bold, underlined) introduced at the same time a restriction site for the enzyme *Apa*LI and removed a restriction site for *Mnl*I, respectively. For A63C and Q125C mutations, restriction sites for the enzymes *Bst*XI and *Xma*I, respectively, were introduced at the 5' end of the forward primers. In the primers

for the S181C mutation, the restriction site for the enzyme *Mnl*I was inserted at the 3' end of the forward primer. In the case of the Q242C mutant, the restriction site for the enzyme *Taq*I was introduced next to the codon for the cysteine mutation. The presence of the mutations was verified by DNA sequencing.
Truncated Cyt *f* was produced in *E. coli* MV1190 (Δ(*lac-pro*AB), *thi*, *sup*E, Δ(*srl-rec*A) 306::Tn10 (tet^r) [F´:*tra*D36, *pro*AB+, *lac*I^{q}ZΔM15]), transformed with pEAF-WT or mutant plasmids and co-transformed with pEC86, containing a cassette for *c*-type cytochrome over-expression.[144] Production and purification of the protein, and spin label attachment of (1-Acetoxy-2,2,5,5-tetramethyl-δ-3-pyrroline-3-methyl) methanethiosulfonate (MTS) or (1-Oxyl-2,2,5,5-tetramethyl-δ-3-pyrroline-3-methyl) methanethiosulfonate (MTSL) were performed as previously reported.[8,29] The yield of protein production ranged from 1 to 2 mg/L of culture. The expression of Q125C mutant was not reproducible and only a small amount of protein was obtained and used for NMR experiments. The concentration of the protein was determined by absorbance spectroscopy using ε_{556} = 31.5 mM^{-1}cm^{-1} for ferrous Cyt *f*.

NMR experiments

All NMR samples contained MES (20 mM, pH 6) and 6% D_2O for lock. The ferric state of Cyt *f* was preserved by addition of $K_3[Fe(CN)_6]$ (50 μM). The pH of the sample was adjusted with small aliquots of HCl (0.5 M) and NaOH (0.5 M). For the chemical shift perturbation experiments Cyt *f* was titrated into Zn-substituted ^{15}N Pc (50 μM). Spectra were recorded at multiple Cyt *f*:Pc molar ratios (0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 2.5, 5.0). Chemical shift perturbation experiments were not performed for Q125C Cyt *f* because of lack of protein. Samples for PRE measurements contained 33 μM Cyt *f* for the Q125C mutant and 66 μM for the other mutants, labeled with either MTS or MTSL. Samples also contained Zn-substituted ^{15}N Pc, 100 μM in the complex with Q125C Cyt *f* and 200 μM for the other Cyt *f* mutants. All NMR spectra were recorded at 298 K on a Bruker Avance III 600 MHz spectrometer equipped with a TCI-Z-GRAD Cryoprobe. The ^{1}H-^{15}N HSQC spectra were acquired with 1024 and 80 complex points in the direct and indirect dimensions, respectively.

NMR data analysis

The NMR spectra were processed with NmrPipe[145] and analyzed with CcpNMR Analysis.[146] Chemical shift perturbation analysis was carried out as describe before.[29]
The PREs were determined according to the procedure described by Battiste and Wagner.[81] The intensity ratio I_p/I_d of the Pc resonances in the presence of MTSL-Cyt *f* (I_p) and MTS-Cyt *f* (I_d) were normalized by dividing them by the average value of the ten largest I_p/I_d values (1.28 for Q7C, 1.63 for Q38C, 1.16 for A63C, 1.13 for N71C and Q104C, 1.37 for Q125C, 0.83 for S181C, 1.06 for S192C and 0.92 for Q242C). The PRE (Γ_2) values were calculated according to the formula:

$$\frac{I_p}{I_d} = \frac{R_{2d}\exp(-\Gamma_2 t)}{R_{2d}+\Gamma_2} \qquad (3.1)$$

R_{2d} represents the transverse relaxation rate in the diamagnetic sample, which was calculated from the linewidth at half height obtained from a Lorentzian peak fit in the direct dimension, by using FuDA (this software was kindly provided by Dr. D. Fleming Hansen, University College London). The symbol *t* indicates the time for transverse relaxation during the pulse sequence (9 ms). The Γ_2 values were extrapolated to the 100% bound state using the experimentally obtained K_D.
The average intermolecular PCS from the ferric haem iron of Cyt *f* to the backbone amide atoms in all Pc conformers was calculated and compared with the experimental PCS previously measured in the wild type complex.[98] The equation used for the PCS calculation, assuming an axial magnetic susceptibility tensor oriented along the Fe-Y1 vector,[46] was:

$$\Delta\delta_{PCS} = \frac{\Delta\chi_{ax}}{12\pi r^3}\left(3\cos^2\theta - 1\right) \qquad (3.2)$$

In which $\Delta\delta_{PCS}$ is the size of the PCS, r is the distance between haem iron and observed Pc nucleus, and θ is the angle between Pc nucleus, haem iron and the nitrogen of the amine group of Y1 in Cyt *f*. $\Delta\chi_{ax}$ is the size of the axial magnetic component of the susceptibility tensor, derived from the g-tensor values measured by EPR spectroscopy on plant Cyt *f* and taken to be 7×10^{-32} m^3, as previously reported for *Nostoc* Cyt *f*.[98] To correct for the possible difference in tensor size for the temperatures of EPR and NMR measurements, 10 K and 298 K, respectively, the $\Delta\chi_{ax}$ was varied from 0.7 to 8.4×10^{-32} m^3.
The agreement between observed (PCS^{obs}) and calculated (PCS^{calc}) PCS was expressed by the PCS Q factor, defined as:

$$Q_{PCS} = \sqrt{\frac{\Sigma\left(PCS^{obs} - PCS^{calc}\right)^2}{\Sigma\left(PCS^{obs} + PCS^{calc}\right)^2}} \qquad (3.3)$$

Monte Carlo simulations of the encounter complex

The structure of the soluble part of Cyt *f* (residues 1-254) used for the calculation was taken from the crystal structure of the b_6f complex from *Nostoc* sp. PCC 7120, PDB entry 2ZT9.[147] The amino acidic sequences of Cyt *f* from *Nostoc* sp. PCC 7120 and sp. PCC 7119 are identical. The structure file for Pc was taken from the PDB entry 2GIM.[42] The hydrogen atoms were added with the module HBUILD[148] of Charmm.[149] To preserve the original structure, only the hydrogen atoms were minimized with the Charmm force field,[150] while the other atoms were kept fixed in their original position. The iron of Cyt *f* and the copper of Pc were considered to be in the oxidized state, like in the experiments. The electrostatic potentials for the single proteins were calculated with APBS.[151] The dielectric constants for Cyt *f* and

the water were set to 4 and 80, respectively. For all electrostatic potentials, a box with a diameter of 225 Å in x, y and z directions, with Cyt *f* centered at the origin of the coordinates frame, was defined. The ionic strength was set to 0.02 M and the temperature to 298 K. The electrostatic potential was calculated with the linearized Poisson-Boltzmann equation.
The docking was performed with the program MC-Dock[57] and was carried out in a similar way as was done before.[28] Cyt *f* was chosen as the receptor and Pc was the ligand to dock. The simulation consisted of 250 runs with 1000000 steps each and was carried out at a temperature of 298 K. Only structures that respected the Metropolis MC criterion[152] were saved resulting in about 2300000 Cyt *f*-Pc orientations. The main difference to the previous simulation consisted in the use of an inclusion grid. The inclusion grid was created by defining a grid with a distance to the surface of Cyt *f* of 3 Å and a grid point separation of 0.5 Å. If any atom of Pc is located within this inclusion grid, the structure was included in the final encounter ensemble, otherwise the orientation was not considered. An ensemble of 5000 Pc orientations, randomly selected, was considered for the calculations. The averaged distances were derived from the ensemble and compared to the experimental distances.

Ensemble docking

Mutations and spin labels were modeled on the structure of Cyt *f* (PDB entry 2ZT9[147]) and four conformations were used to represent the mobility of the spin label.[77] The structure of Pc was taken from PDB entry 2GIM.[42]
The Γ_2 were converted into distances for structure calculations using Equation 3.4:

$$r = \sqrt[6]{\frac{\gamma^2 g^2 \beta^2}{20\Gamma_2}\left(4\tau_c + \frac{3\tau_c}{1+\omega_h^2\tau_c^2}\right)} \qquad (3.4)$$

Where *r* is the distance between the oxygen atom of MTSL and the Pc amide proton, γ is the proton gyromagnetic ratio, *g* is the electronic g-factor, β is the Bohr magneton, ω_h is the Larmor frequency of the proton and τ_c is the rotational correlation time of the MTSL oxygen-proton vector. τ_c was taken to be 30 ns on the basis of the HYDRONMR[153] prediction of the rotational correlation time for the Pc-Cyt *f* complex.
The restraints for the calculations were obtained according to Equation 3.5:

$$\Gamma_2^{obs} = f_1\Gamma_2^{ens} + f_2\Gamma_2^{final} \qquad f_1 + f_2 = 1 \qquad (3.5)$$

The ensemble Γ_2 (Γ_2^{ens}) was calculated as the difference between observed Γ_2 (Γ_2^{obs}) and back calculated Γ_2 from the model of the final complex (model 1, PDB entry 1TU2) (Γ_2^{final}). The calculations were carried out with f_2 values of = 0, 0.15, 0.25, 0.35, 0.5, 0.65, 0.75, 0.85, 0.95 and 1.0. The restraints were grouped into three classes as described before.[29]
A description of the encounter complex was obtained with restrained rigid-body

docking in Xplor-NIH 2.9.9[154] to minimize the difference between observed and back calculated distances for all spin labels. Calculations were carried out using either a single Pc conformer or an ensemble of Pc molecules, with between 2 and 20 copies. The distances (r) between an amide proton and the oxygen atom of MTSL were r^{-6} averaged for all MTSL orientations and all Pc conformers. Cyt *f* and Pc were both considered as rigid bodies, the coordinates of Cyt *f* were fixed, and Pc ensemble members were allowed to move individually in a restrained rigid-body molecular dynamics calculation. Overlap of Pc copies was allowed, since the ensemble represents a distribution of states. Similarly, overlap of MTSL conformers with other MTSL or Pc copies was allowed. For the visualization of the final encounter complex ensemble 150 dockings were performed, yielding 144 ensembles of seven Pc conformers, with a difference in the total restraint energy ≤20%.

The ensembles from separate dockings were evaluated by calculating the average violation over all experimental distances. Class 1 and class 3 restraints are not easily expressed in a Q value. Violations provide a better representation of the fit of all three classes of restraints. Class 2 violations were defined as the absolute difference between experimental and calculated distances for a certain amide nucleus. Class 1 and class 3 restraints were defined as that difference only for back-predicted distances that were above 14 Å and below 23 Å, respectively. The ensemble violation is the average violation for all residues and all spin labels.

Results

Introduction of paramagnetic probes on Cyt *f*

To determine the extent of surface area of Cyt *f* being sampled by Pc in the encounter complex, nine cysteine mutants of Cyt *f* were made for the attachment of nitroxy spin labels. The Cyt *f* mutants were created for the positions Q7, Q38, A63, N71, Q104, Q125, S181, S192 and Q242. The mutation sites Q7, A63, N71, Q104 and S192 are near to the Pc binding site indicated by the solution model,[98] whereas the remaining four mutations are located elsewhere (Figure 3.1, *central panel*). To preserve the original electrostatic potential of Cyt *f*, only polar, uncharged amino acids and one Ala were selected for mutation to cysteine. ^{15}N Pc was produced with copper and substituted by Zn^{II} to eliminate the paramagnetic effect and possible interference of the ET reaction caused by the presence of Cu^{II}.[154]

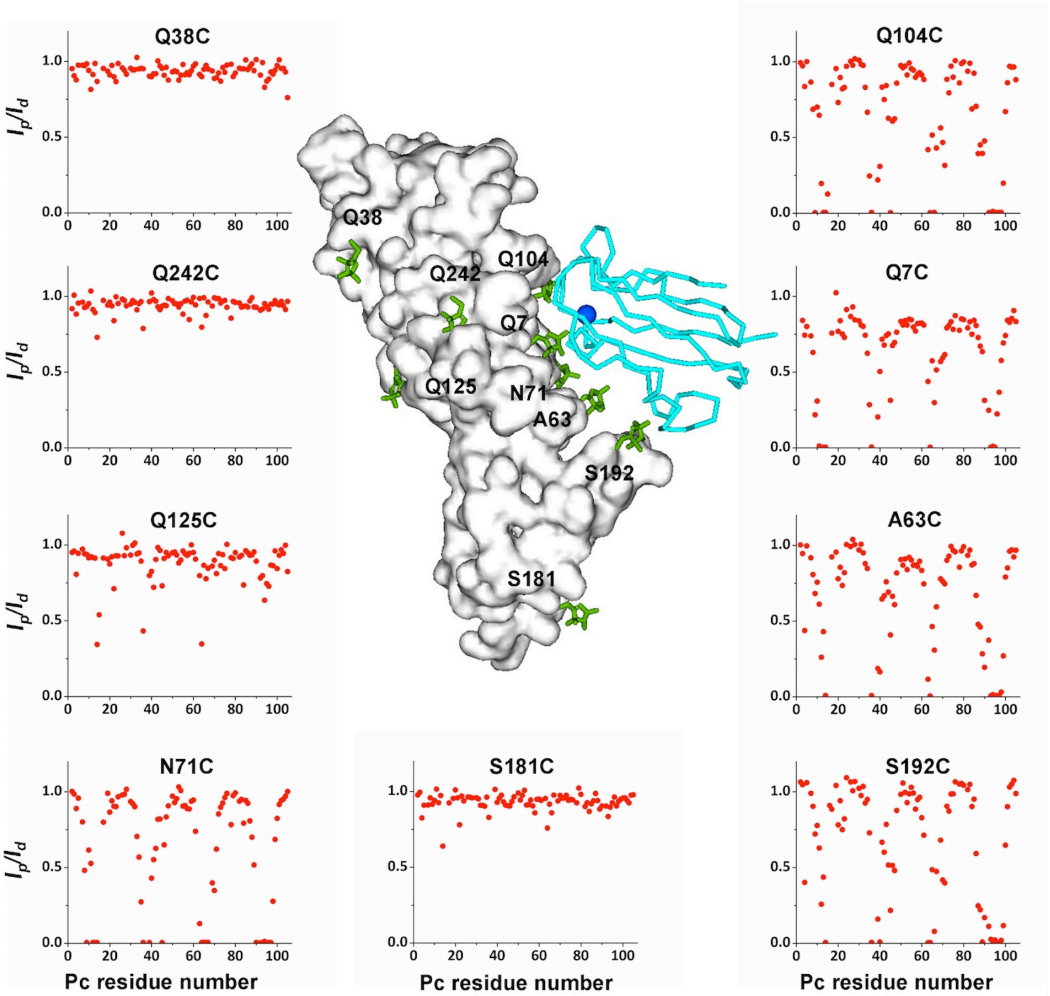

Figure 3.1. PRE in the Pc-Cyt *f* complex. *Central panel*. Location of the spin labels (green sticks) modeled on the *Nostoc* sp. PCC 7119 Pc-Cyt *f* complex (PDB entry 1TU2, model 1[98]). Pc is shown in cyan C_α trace and the copper as a blue sphere. Cyt *f* is shown as white surface. *Side panels*. The I_p/I_d ratios (red dots) are plotted against the Pc residue number for each of the spin label position on Cyt *f*.

To test whether the presence of spin label interferes with the Pc-Cyt *f* binding, chemical shift perturbation (CSP) analysis was performed for Pc bound to Cyt *f* wild type and mutants conjugated to the diamagnetic control label MTS. Cyt *f* was thus titrated into a solution of ^{15}N Zn-Pc and HSQC spectra were acquired at each titration point. The CSP curves for the most affected residues were fitted to obtain a dissociation constant for each complex (Figure 3.2).

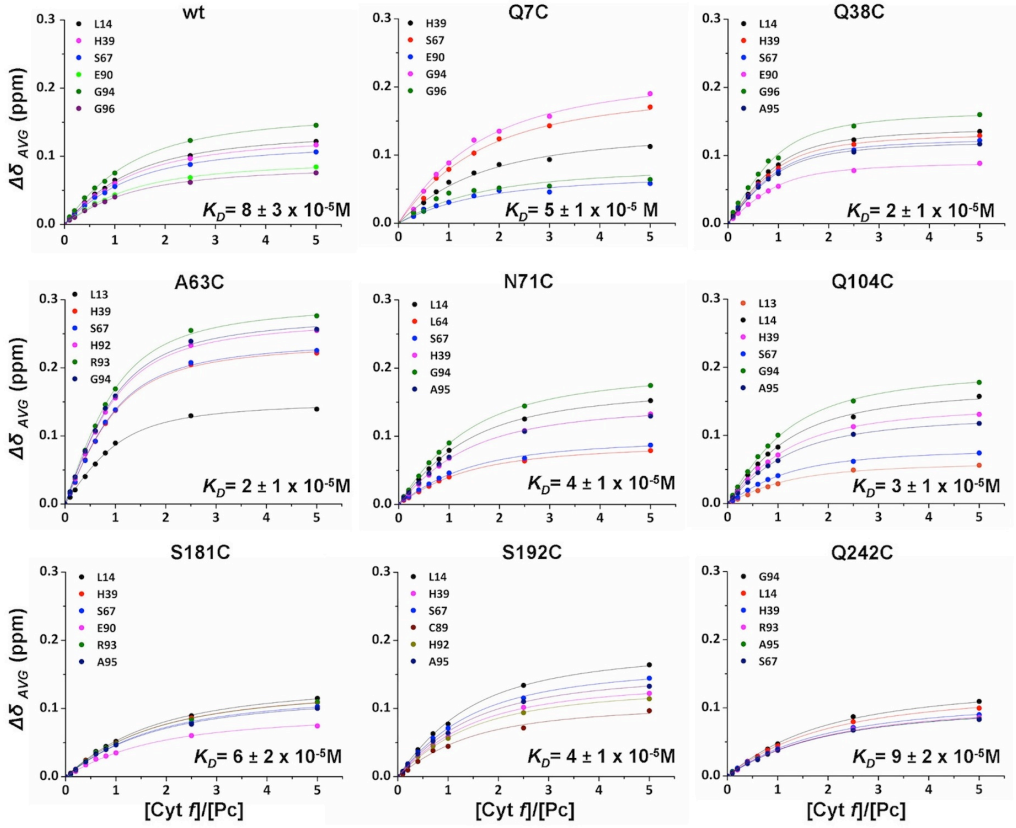

Figure 3.2. The interaction of *Nostoc* Zn-substituted Pc with wild type Cyt *f* and MTS-conjugated variants. The binding curves for selected residues were fitted globally to a 1:1 binding model.

The K_D values are also listed in Table 3.2.

Table 3.2. Dissociation constants of the complexes formed by *Nostoc* Zn-Pc with wild type and MTS-conjugated Cyt *f*. The errors are indicated in parentheses and represent the precision of the fit.

Cyt *f* mutant	K_D (10^{-5} M)
wild type	8 (3)
Q7C-MTS	5 (1)
Q38C-MTS	2 (1)
A63C-MTS	2 (1)
N71C-MTS	4 (1)
Q104C-MTS	3 (1)
S181C-MTS	6 (2)
S192C-MTS	4 (1)
Q242C-MTS	9 (2)

The K_D value for the wt complex of 8 ± (3) × 10^{-5} M is similar to the reported values of 4 × 10^{-5} M for Cu-Pc[49] and 6 × 10^{-5} M for Cd-substituted Pc.[49]

Most of Cyt *f* variants yielded K_D values within the experimental error of that of the wild type. Moreover, the binding maps, obtained by coloring the protein residues according to the size of CSP, present a pattern similar to the wild type, indicating that the mutations and the attachment of MTS at these positions cause no significant effects on the affinity of Pc for Cyt *f* and orientation of Pc with the respect to Cyt *f* in the complex. In each case Pc binds predominantly via the hydrophobic patch and the region around Arg 93 (Figure 3.3), similarly to previously reported data on Cd-substituted Pc in the presence of reduced Cyt *f*.[98]

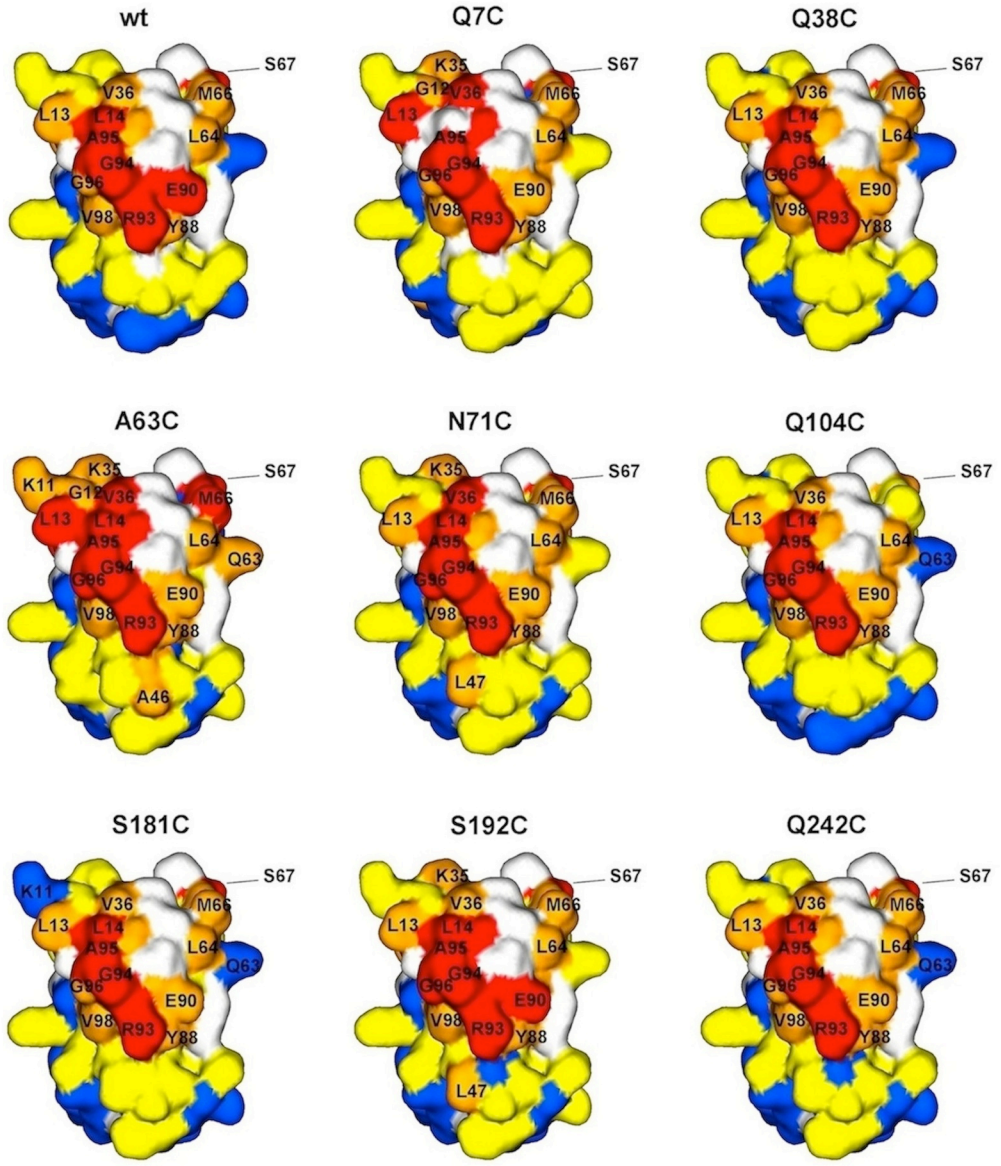

Figure 3.3. Chemical shift perturbation maps of *Nostoc* Zn-substituted Pc in the presence of wild type and MTS-conjugated Cyt *f*, colour-coded on a surface model of Pc (PDB entry 2GIM), with red, $\Delta\delta_{avg} \geq 0.10$ ppm; orange, $\Delta\delta_{avg} \geq 0.05$ ppm; yellow, $\Delta\delta_{avg} \geq 0.02$ ppm; blue, $\Delta\delta_{avg} < 0.02$ ppm. Prolines and residues with overlapping resonances are in white.

The Q38C-MTS and A63C-MTS Cyt *f* variants exhibited K_D values of 2 ± (1) × 10^{-5} M, which represent a small increase of binding affinity. In the case of Q38C-MTS Cyt *f*, the average size of the CSP and binding map were similar to wild type. Larger perturbations of the resonance positions were observed for binding of Pc to Cyt *f* A63C-MTS than in the other studied cases. The largest CSP were about twice as large as those in the presence of wild type Cyt *f* (Figure 3.4A). Interestingly, the binding map is still similar to that of wt, although the effects of binding are stronger (Figure 3.4B). Under the assumption that CSPs predominantly represent the final state,[31,72] this observation suggests that Pc binds Cyt *f* A63C-MTS in the same orientation as wt Cyt *f* but that the final state is more populated and the encounter state less.

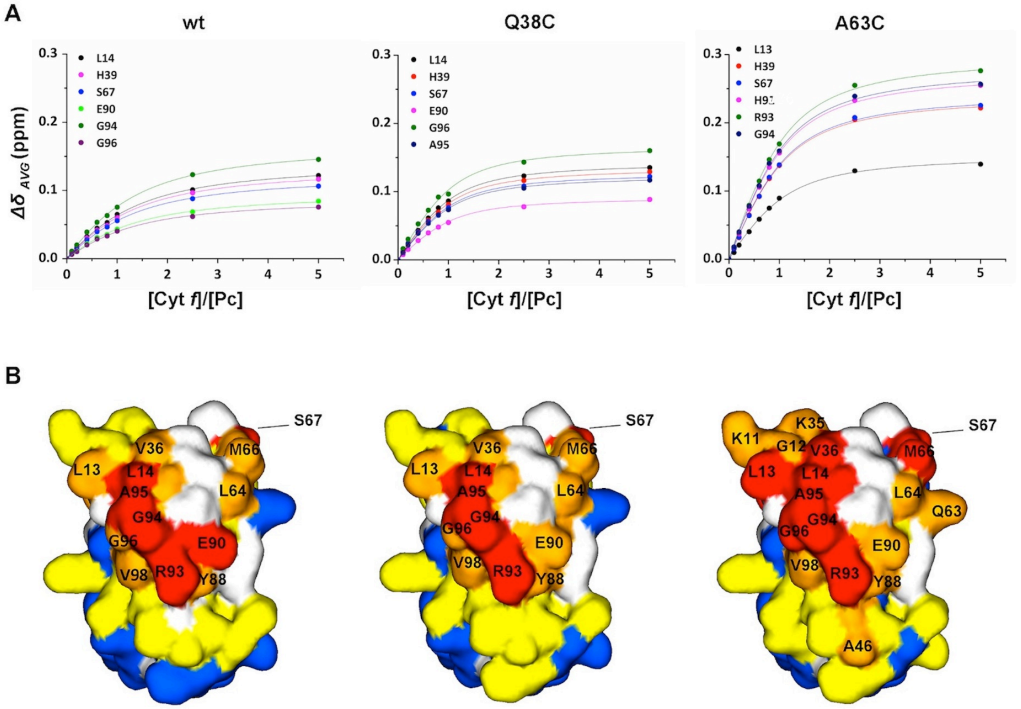

Figure 3.4. The interaction of *Nostoc* Zn-substituted Pc with wild type Cyt *f* and Q38C and A63C MTS-conjugated variants. A) Binding curves for selected residues were fitted globally to a 1:1 binding model.[29] B) Chemical shift perturbation maps color-coded on a surface model of Pc (PDB entry 2GIM[42]), with red, $\Delta\delta_{avg} \geq 0.10$ ppm; orange, $\Delta\delta_{avg} \geq 0.05$ ppm; yellow, $\Delta\delta_{avg} \geq 0.02$ ppm; blue, $\Delta\delta_{avg} < 0.02$ ppm. Prolines and residues with overlapping resonances are in white.

Paramagnetic relaxation enhancements

To determine whether PREs could arise from unspecific interactions, free MTS (diamagnetic) and MTSL (paramagnetic), with a concentration corresponding to that of spin labeled Cyt *f* (66 μM, see below), were mixed with ^{15}N-labeled Zn-Pc (200 μM). No line broadening of the resonances was observed in the presence of the paramagnetic spin label, indicating that unspecific interactions with the label are not significant under these conditions.

Then, MTSL was attached to each of the nine Cyt *f* mutants and the tagged proteins were titrated to Pc to a molar ratio of 1:0.3 for Pc:Cyt *f*. At this ratio, the average fraction of Pc bound to Cyt *f* is 24%. The CSP studies indicated that association and dissociation are in the fast exchange regime, so the observed PREs are a weighted average of free Pc, encounter complex and final complex. Thus, the PREs can be extrapolated to the 100% bound state (encounter state + final complex) by dividing by the fraction of bound Pc.

Spin labels attached to Cyt *f* near the binding site for Pc in the final complex, namely at the positions Q7, A63, N71, Q104 and S192, caused an extensive broadening of Pc resonances, reflected in a decrease of the I_p/I_d ratio, the ratio of peak intensities in the spectra of the paramagnetic and diamagnetic samples (Figure 3.1). Surprisingly, mutants with the spin label located on the backside of Cyt *f* relative to the binding site also yielded moderate to strong PREs. For three of these mutants, Q38C, S181C and Q242C, only moderate effects were observed, whereas the spin label at position 125 caused strong PRE on two Pc residues, L14 and L64, which are part of the hydrophobic patch. From the I_p/I_d ratios, the PRE (Γ_2) were determined and extrapolated to 100% bound Pc. The PREs were mapped on the surface of Pc, the front side of the protein is shown in Figure 3.5.

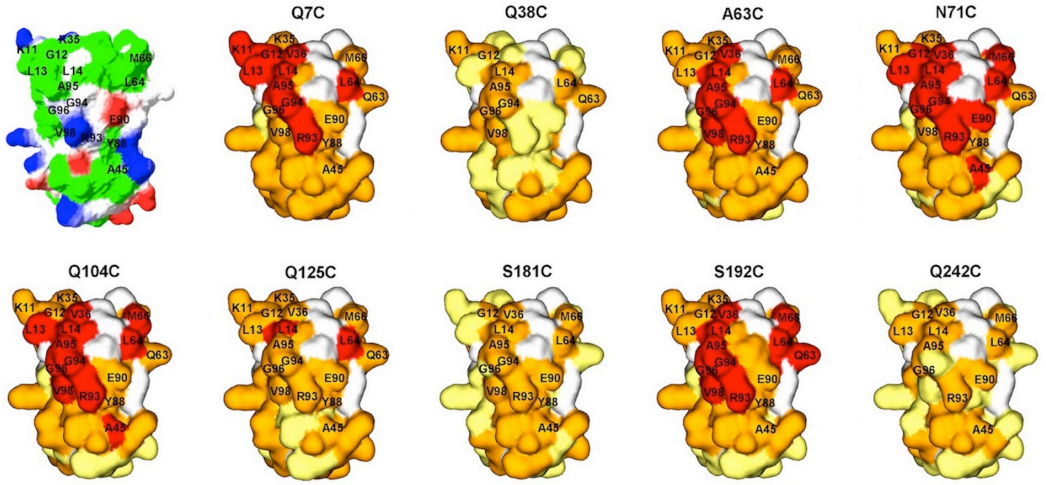

Figure 3.5. PRE maps of Zn-substituted Pc bound to MTSL-conjugated Cyt *f*, color-coded on a surface model of the front side of Pc (PDB-entry 2GIM). The sites of spin label attachment are indicated in Figure 3.1, central panel. Red, $\Gamma_2 \geq 200$ s^{-1}; orange, 10 s$^{-1} < \Gamma_2 < 200$ s^{-1} and yellow $\Gamma_2 \leq 10$ s^{-1}. Prolines and residues with overlapping resonances are white. Top left, the charge distribution of Pc with negatively and positively charged side chains shown in red and blue, respectively. Hydrophobic side chains are shown in green, polar side chains are in white.

The PRE patterns observed in the presence of spin label attached near the main binding site are very similar. This result is surprising, because the labels are located on different sides of Cyt *f* in the structure of the final complex as shown in Figure 3.1, and thus it is expected that different regions of Pc would be affected. The results suggest that Pc samples an extensive area of Cyt *f* predominantly with one face oriented toward it, since no strong PREs were observed on the back side of Pc (Figure 3.6).

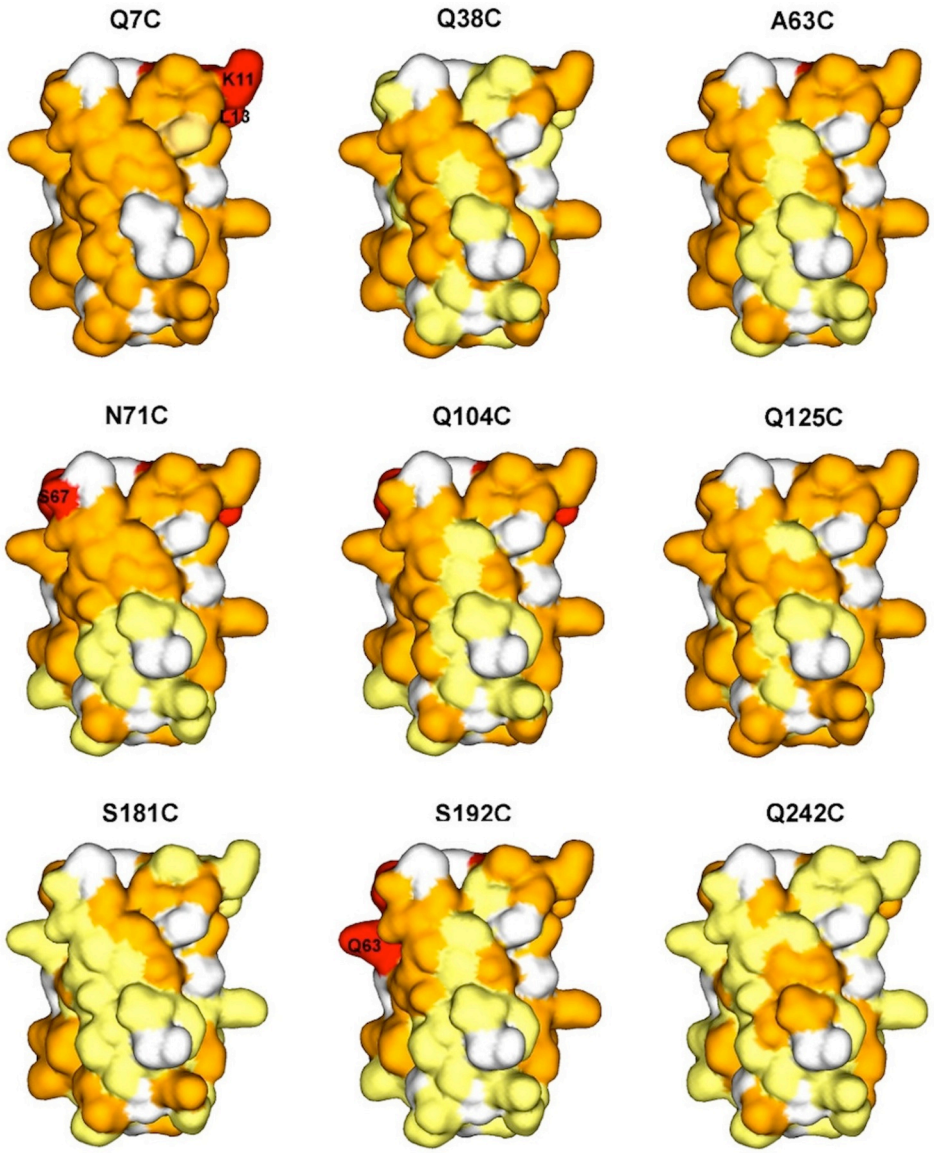

Figure 3.6. PRE maps of Zn-substituted Pc bound to MTSL-conjugated Cyt *f*, color-coded on a surface model of the back side of Pc (PDB entry 2GIM), the sites of spin label attachment are indicated in Figure 3.1, central panel. Red, $\Gamma_2 \geq 200$ s^{-1}; orange, 10 s$^{-1} < \Gamma_2 < 200$ s^{-1} and yellow $\Gamma_2 \leq 10$ s^{-1}. Prolines and residues with overlapping resonances are white.

The maps also resemble the CSP maps in Figures 3.3 and 3.4, confirming that the hydrophobic patch and the region around Arg 93 are the interaction sites of Pc in the complex with Cyt *f*. The comparison of the PRE maps with the charge distribution map of Pc (Figure 3.5, top left) indicates that among the residues, which experience most PRE, only the minority is negatively (E90) or positively (K11, K35 and R93) charged, whereas the majority has a hydrophobic nature. Residues L13, L14, V36, L64, A95 and V98 are part of the hydrophobic patch, which also represents the main binding site and the likely site for ET.[98]

Monte Carlo simulations

Visualization of the encounter state on the basis of the PRE data is not straightforward, because the data represent a weighted average of all orientations of Pc within the complex, and thus, an infinite number of ensembles can produce the experimental data set. The encounter complex of Cyt c and CcP was successfully visualized by combining PRE data and rigid-body MC simulations,[28] showing that the formation of this encounter complex is solely driven by electrostatic interactions. In MC docking, a mobile protein is docked to a target molecule under the influence of an electrostatic field and MC sampling.[57] In this way, charge-charge interactions represent the only force that brings together the proteins. Following the same rationale, MC simulations for Pc-Cyt *f* complex were performed and the Boltzmann distribution of orientations of Pc in complex with Cyt *f*, and vice versa, were obtained. The centers-of-mass of Pc (Figure 3.7A) and Cyt *f* (Figure 3.7B) are shown as blue and green spheres, respectively, around the interaction partner, shown as surface model.

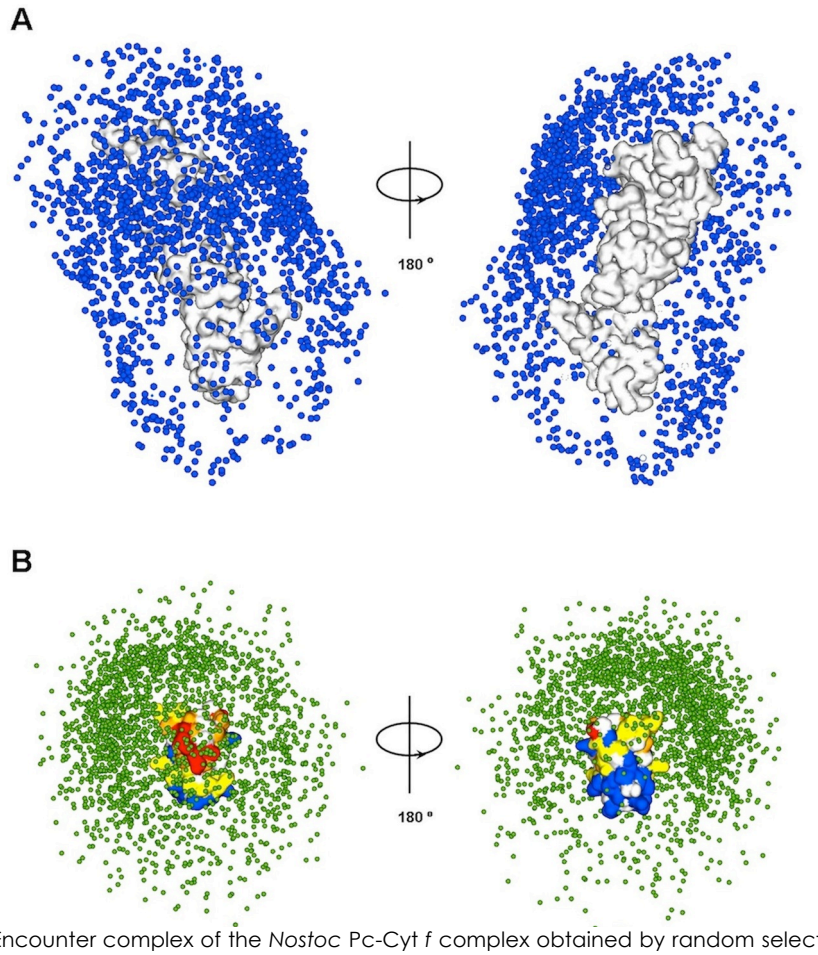

Figure 3.7. Encounter complex of the *Nostoc* Pc-Cyt *f* complex obtained by random selection of 2000 structures from the MC simulations. A) Cyt *f* is shown as a white surface and Pc centers-of-mass are represented by blue spheres. B) Pc is shown as a surface color-coded according to the CSP in the presence of wild type Cyt *f* and Cyt *f* centers-of-mass are represented by green spheres.

In the MC ensemble Pc is widely spread over the surface of Cyt *f* in correspondence with the negative charges distribution. Cyt *f* is over-all negative with most charges in the region surrounding the haem on the large domain and lower charge density on the surface opposite of the haem. These results are inconsistent with the PREs observed in the presence of spin labels located at positions far from the haem, such as Q125 (backside) and S181C (small domain). The distribution of Cyt *f* around Pc is off-center from the CSP map obtained for binding to Cyt *f*. Thus, qualitatively the MC ensembles are not in complete agreement with the experimental data.

The observed PREs result from the contributions of both the encounter and the final complex, whereas the MC ensemble is assumed to represent mostly the encounter state. To separate the PRE contributions of the two states, the PREs from the solution model of the final complex were back calculated and subtracted from experimental PREs assuming a population of the final state (f_2) varying from 0 – 1. The resulting PREs represent the encounter state at decreasing population and these were converted into distances and compared with the average distances calculated from the MC ensemble (See Experimental Section for details). Ensembles composed of the 100, 1000, 2500 and 5000 randomly selected structures were considered for the analysis. Independent of the size of the MC ensemble and of the population of the two states, no good match with experimental data was found (Figure 3.8).

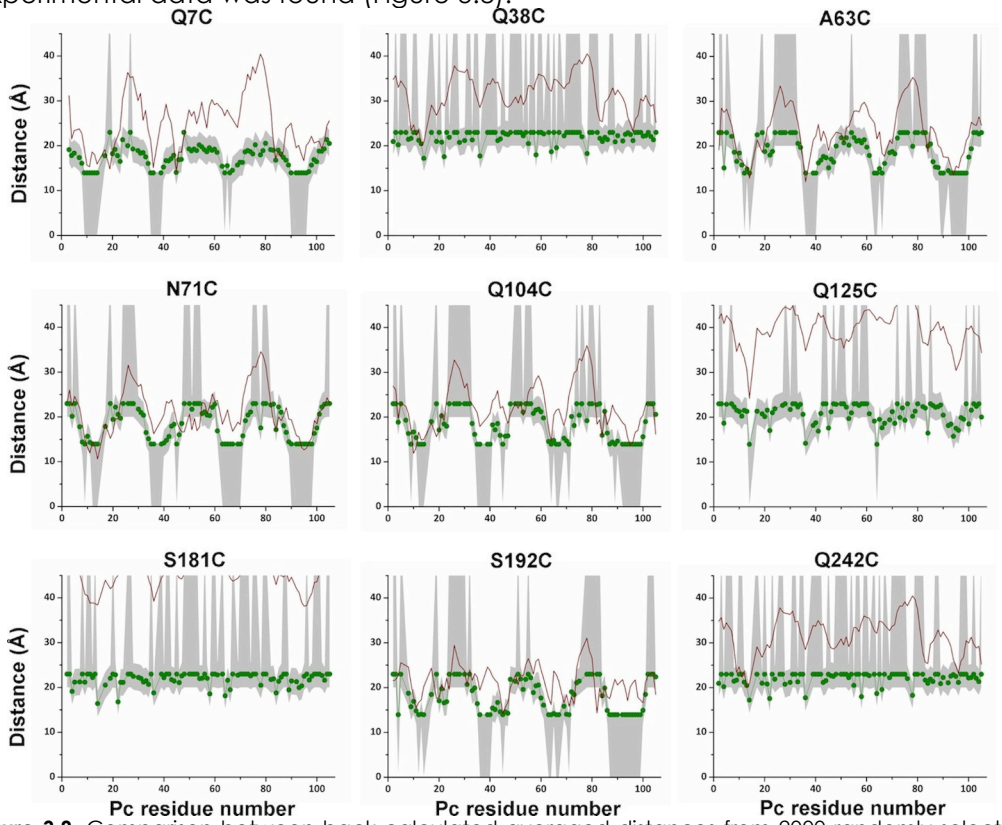

Figure 3.8. Comparison between back calculated averaged distances from 2000 randomly selected structures of the MC simulation (red line) assuming f_1= 1 and the experimental distances (green circles and lines). The grey areas indicate the error margins of experimental data.

These findings suggest that the formation of the Pc-Cyt *f* encounter complex is not exclusively driven by electrostatic forces. Other contributions must play a significant role and, therefore, MC simulations cannot provide a complete description of this encounter complex.

Ensemble docking

The quantitative interpretation of PRE for the visualization of transient encounter complexes requires the use of PRE restraints in docking calculations.[84] An ensemble of orientations that represents the encounter state and agrees with the data, is generated by docking several conformers of a protein (Pc in this work) simultaneously while minimizing the difference between the back calculated PRE averaged over all conformers and the experimental data. This procedure is repeated many times, and because many different ensembles fit the experimental data, the result is a 'cloud' of orientations.[140,156-161] An ensemble of non-interacting Pc structures was generated and docking calculations were performed with PRE restraints arising from all nine spin labels at the same time. Though variant Cyt *f* A63C-MTSL appeared to have some influence on the equilibrium between encounter state and final complex (see above), it was included in the calculations, because we found that the description of the encounter complex was similar, whether or not these restraints were included in the calculations. In our hands, converting PREs to distances worked best, probably because Pc approaches some spin labels closely. Due to inverse sixth power distance dependence, very small movements at short distance result in very large PRE changes that skew the outcome of the calculations. A repulsion function to avoid steric collision between Cyt *f* and the Pc molecules was the only other interaction included in the calculations. Calculations were performed by varying the size (*N*) of the docked ensemble, which ranged from 1 to 20 copies of Pc. The generated ensembles were evaluated by calculating the average violation over all experimental distances (See Experimental Section). As can be seen in Figure 3.9A large decrease of the average violation was observed, going from a single copy of Pc up to $N = 7$, while further increase of the number of Pc molecules ($N > 7$) did not improve the fitting.

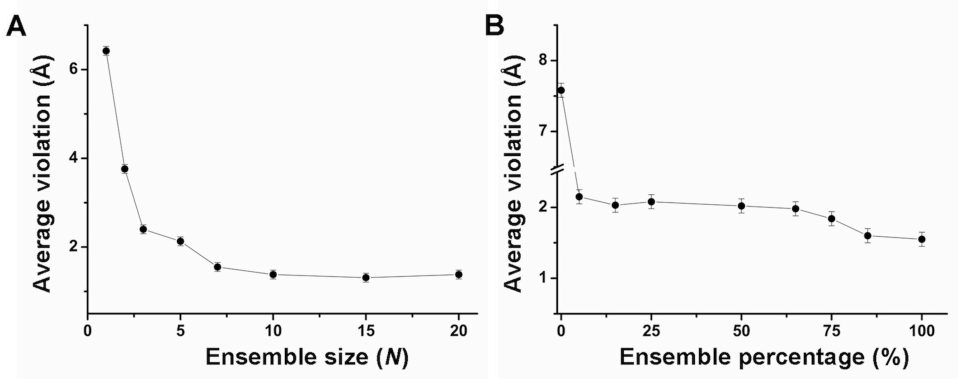

Figure 3.9. Plot of the average violation of all experimental distances versus the number (*N*) of Pc copies used in ensemble docking (A) and versus the ensemble percentage included in the restraints for the calculations (B). Error bars represent 2 × SD of the average violations obtained from three independent calculations performed with N= 1+7 and an encounter percentage of 50% (f_1= 0.5).

Thus, a combined docking of seven Pc copies simultaneously can produce a population distribution that can mostly satisfy the experimental PREs.

The solution structure of the final complex was previously reported,[98] taking advantage of intermolecular PCS generated by the paramagnetic Fe^{III} of Cyt *f* on Pc nuclei. However, the results in Figure 3.9A show that a single orientation cannot account for the PRE data. In a earlier study we showed that the model of the final complex is already insufficient to explain the PRE data of only three nearby spin labels.[29] Thus, the PRE data describe a combination of the final state and encounter complex. Following this rationale, calculations were carried out with N = 1+7, where 1 represents the final complex and 7 the number of copies in the ensemble.[158] The contribution of the final complex to the experimental PREs was subtracted to obtain the PREs of the encounter state only, in analogy to what was done for the MC calculations. Thus, the back calculated PREs from the solution model[98] were subtracted from the experimental values, assuming a population of the final complex (f_2) between 0 and 1. The resulting PREs were converted to distances and used for ensemble docking. Figure 3.9B presents the average violation as a function of the fraction of encounter state (f_1). The violations unequivocally indicate that the measured PREs do not derive from the final structure alone. Interestingly, a small fraction of encounter state is sufficient to decrease the average violation sharply. The average violation decreases slightly from 2.15 for f_1 = 0.05 encounter complex to 1.55 for f_1 =1. In Figure 3.10 the results from the simulation carried out using seven conformers of Pc (N= 7) and assuming a pure encounter state (f_1 = 1) are shown.

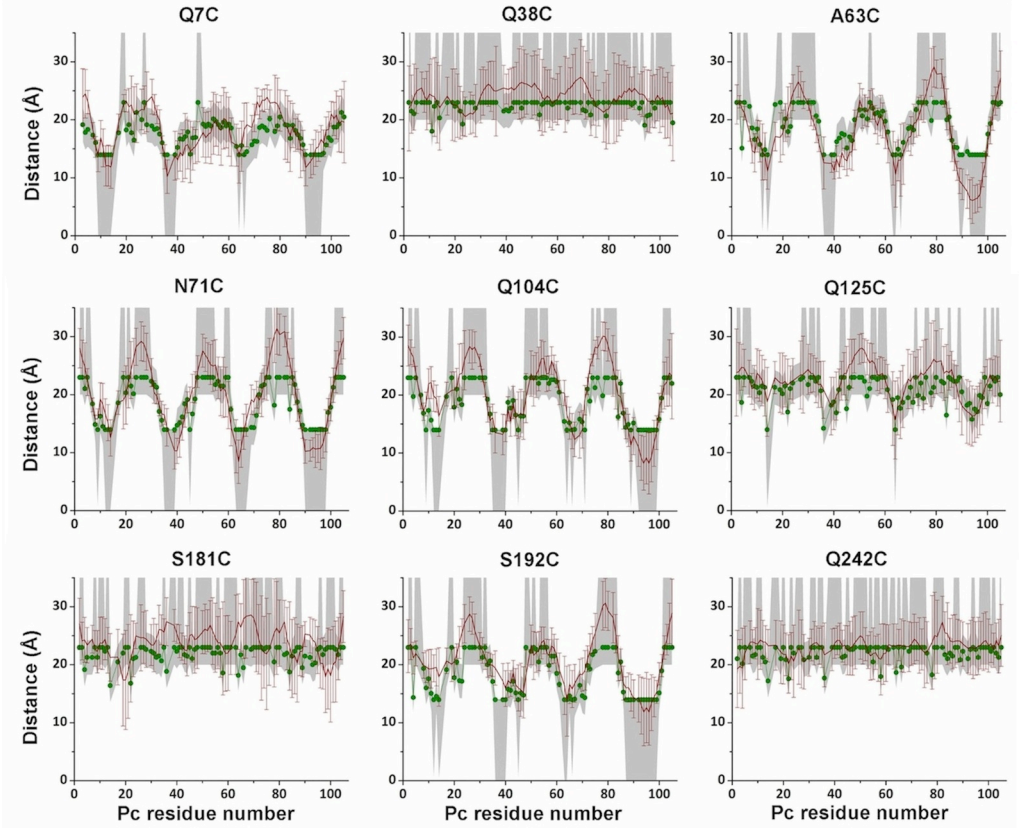

Figure 3.10. Ensemble docking. Experimental and back calculated average distances between Pc amide protons and oxygen atoms of MTSL conjugated to Cyt *f* are plotted against the Pc residue number. The green circles and lines represent the experimental distances and the grey areas indicate the error margins. The average distances back calculated from twenty ensembles are shown as a red line with error bars representing the SD. Calculations were performed with N= 7 and f_1= 1.0.

The ensemble structures were used to back-calculate the distances between the oxygen atom of MTSL and Pc amide protons (red line) and these were compared with the experimental distances (green circles and line). The generated encounter complex fits the experimental data well, being within the error margins for most residues, although small deviations are observed for some residues at several spin label positions. The considerable standard deviations observed for the distances for the generated ensembles (error bars for red line) are noteworthy, because it is a clear illustration that rather different ensembles of seven Pc copies can fit the large experimental data set equally well, emphasizing the nature of the 'inverse problem' mentioned above.

Estimation of the fraction of the encounter complex

The structure of the final complex was based on experimental PCS, not on PRE, and consists of a single orientation of Pc relative to Cyt *f*, so by reducing the contribution of the final complex, it is expected that it is easier to create an ensemble that matches the experimental PREs. Therefore, the small decrease of the average violation with increasing fraction of the encounter complex (Figure 3.10B) may not be significant, indicating that the PRE data cannot distinguish between a fraction of the encounter complex of 5% and 100%. Since both PCS and PRE account for minor species present in solution, PCS were back calculated for the generated encounter complexes and compared with the experimental PCS data. To correlate experimental and back calculated PCS, a Q factor (equation 3.3) was calculated for different fractions of final structure (f_2) (Figure 3.11).

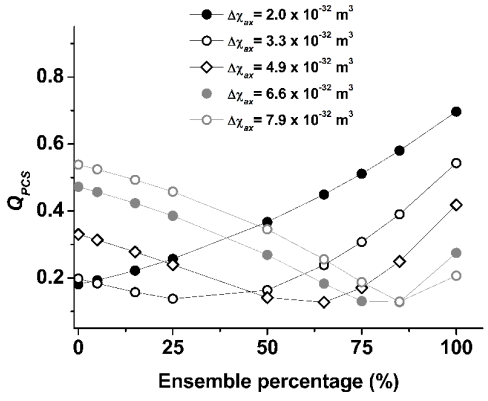

Figure 3.11. Q factors calculated for a combination of experimental PCS measured for the specific complex and back calculated PCS from the encounter complex obtained at different percentage of the encounter complex. The Q factors were calculated for different values of a scaling factor for the size of the axial component of the magnetic susceptibility tensor.

The size of the axial component of the magnetic susceptibility anisotropy ($\Delta\chi_{ax}$) of Cyt *f* Fe^{III} is not known precisely, so a range of values was tested. For a final complex only (f_1 = 0), the best fit of the PCS is found for $\Delta\chi_{ax}$ values less than 50% of the one derived from EPR data, in line with earlier findings.[98] The low-lying excited states for a low-spin ferric haem explain why the g tensor at 10 K cannot readily be used to calculate the $\Delta\chi_{ax}$ at 298 K. For all but

very low values of $\Delta\chi_{ax}$, the combination of final complex and encounter ensemble (determined using the PRE ensemble docking) improves the fit between experimental and back calculated PCS. Unfortunately, the lowest Q-value achievable is always about 0.1 (see Figure 3.11), and this minimum is found at increasing values of f_1 for larger $\Delta\chi_{ax}$ values. Therefore, establishing the encounter state fraction on the basis of PCS is not possible as long as $\Delta\chi_{ax}$ cannot be established. For comparison, $\Delta\chi_{ax}$ for Cyt *c* is about 3.3×10^{-32} m^3.[74] If $\Delta\chi_{ax}$ of Cyt *f* would be the same, the fraction f_1 would be 0.25.

Visualization of the encounter complex

To represent the encounter complex an ensemble of 144 solutions, for a total of 1008 Pc molecules was generated (N=7; f_2= 0). As shown in Figure 3.13 Pc visits a large area of Cyt *f*. The density plot (Figure 3.12A) shows the centers-of-mass of Pc colored according to the density, with red and blue representing the most and least populated positions, respectively.

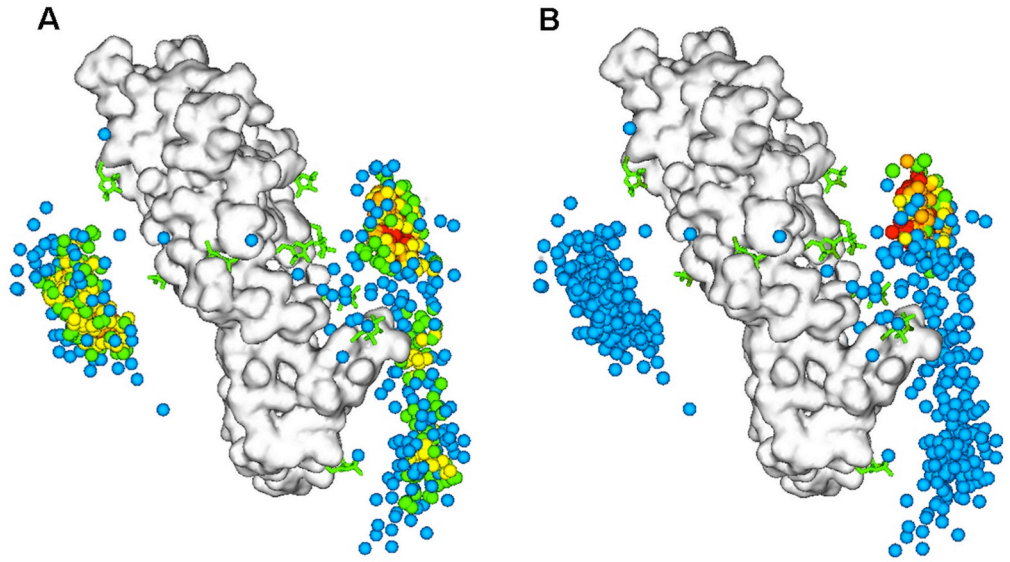

Figure 3.12. Encounter complex of the Pc-Cyt *f* complex. Cyt *f* is shown as a white surface and spin labels as green sticks. Pc centers-of-mass are represented by spheres. In panel A, Pc centers of mass are color-coded to indicate the density of the distributions, decreasing from red to blue. Densities were determined by counting the number of neighbors within 2.5 Å. In panel B, Pc centers-of-mass are color-coded to indicate the distance between Cu in Pc and Fe in Cyt *f*, increasing from red to blue (red ≤ 16 Å; orange ≤ 18 Å; yellow ≤ 20 Å; green ≤ 22 Å; blue > 22 Å).

Two defined encounter locations can be distinguished. The first one is close to the binding site in the final complex, and the second on the other side of Cyt *f*, opposite to the binding site in the final complex. These locations are discussed further below. It should be noted that even with nine spin labels it was not possible to sample the encounter state at all locations on the Cyt *f* surface sufficiently. The cytochrome is a very elongated protein and from the spin label positions in Figure 3.1 (central panel), the regions that were not sampled can be identified. It

cannot be excluded that the area sampled in the encounter complex is still larger than the surface area covered in our experiments. Nevertheless, from the current analysis, it is obvious that Pc samples quite a significant fraction of the surface of its partner.
It is thought that in ET systems the formation of the encounter complex reduces the dimensionality of the search for the active site and increases the probability of ET.[21] To determine which of the encounter complex orientations were compatible with rapid ET, the distances between Cu, in Pc, and Fe, in Cyt *f* were calculated and the centers-of-mass of Pc were color-coded accordingly (Figure 3.12B). Structures with Cu-Fe distance ≤ 16 Å (red dots), thus in principle suitable for fast ET,[162] are exclusively located in front of the haem, in the vicinity of the final structure.

Discussion

The structure of Pc-Cyt *f* final complex from the cyanobacterium *Nostoc* sp. PCC7119 was solved by NMR on the basis of PCS data.[98] The results were consistent with a conformation that accounts for the ET reaction between the two proteins. A recent PRE study[29] on the same system demonstrated that the complex exists partly in a dynamic ensemble of orientations. In this study we characterized the encounter complex of Pc and Cyt *f* by attaching the spin label MTSL to nine sites on Cyt *f*, one at a time, and measuring PRE of the backbone amide protons of Pc. MTSL is a small hydrophobic molecule and its presence in particular region of the protein could influence the complex formation with the interaction partner. Some interference by the spin label in the Pc-Cyt *f* complex formation was observed only at the position A63, but exclusion of the data of this spin label had little influence on the results. PREs were mainly observed for spin labels situated near to the binding site indicated by the PCS based model, although effects were also measured for the remaining spin label positions. The detection of widespread PREs clearly indicates that Pc samples a large surface area. The similarity of PRE patterns observed in the presence of spin labels close to the binding site (Figure 3.5) suggests that Pc is approaching Cyt *f* with the residues forming the hydrophobic patch and the region around R93. Independent of the location of the paramagnetic probes, the residues that experience the strongest PREs reside in these regions, implying that Pc is always oriented in the same way toward Cyt *f*.
The general model of protein association states that the formation of the encounter complex is an electrostatically driven process.[21] On the basis of this assumption, computational approaches, such as BD and MC docking, have been developed to describe the encounter complex as end-point of electrostatic steering towards the interacting partners.[57,99] These methods successfully described the encounter complex in cases of protein association guided by charge-charge interactions.[28,99,100,109,137,163] The MC simulations of the Pc-Cyt *f* encounter complex did not produce a result in accordance with the experimental data. It is important to note that despite the overall electrostatic potential of Pc and Cyt *f* being positive and negative, respectively, Pc shows a

defined charge distribution, but Cyt *f* has a diffuse surface charge. For interaction partners with weak charge complementarity, it has been demonstrated that electrostatics do not play an exclusive role in protein association, but desolvation is the main driving force in binding.[138] Furthermore, the interface of the encounter complex can bury a significant solvent-accessible surface area, suggesting a role for hydrophobic interactions in the stabilization of the encounter complex.[27] In the Pc-Cyt *f* complex from the cyanobacterium *Ph. laminosum*, the association is dominated by hydrophobic interactions and only hydrophobic contacts stabilize the final complex.[54] In the *Nostoc* Pc-Cyt *f* system, the specific binding interface and the putative ET sites on both proteins comprises hydrophobic regions,[98] similar in size and composition to the *Ph. laminosum* counterparts, but electrostatic forces play a significant role in the association reaction[8,66] and in the orientation of the final complex.[49] Since already in the encounter complex Pc is oriented toward Cyt *f* with its hydrophobic patch, we propose that during the initial stage of the encounter complex formation, long-range electrostatics pre-orient Pc towards Cyt *f* and hydrophobic interactions keep Pc close to the surface of Cyt *f* and help to stabilize the encounter state.

The simulation of the encounter complex on the basis of the experimental PREs resulted in two distinct encounters. One is located at the side of the specific binding surface of Cyt *f*, the other one at the opposite side. In Figure 3.13 the encounter complex is superimposed with the b_6f complex as it is found in the thylakoid membrane, with the Pc centers-of-mass colored on the bases of the Cu-Fe distance.

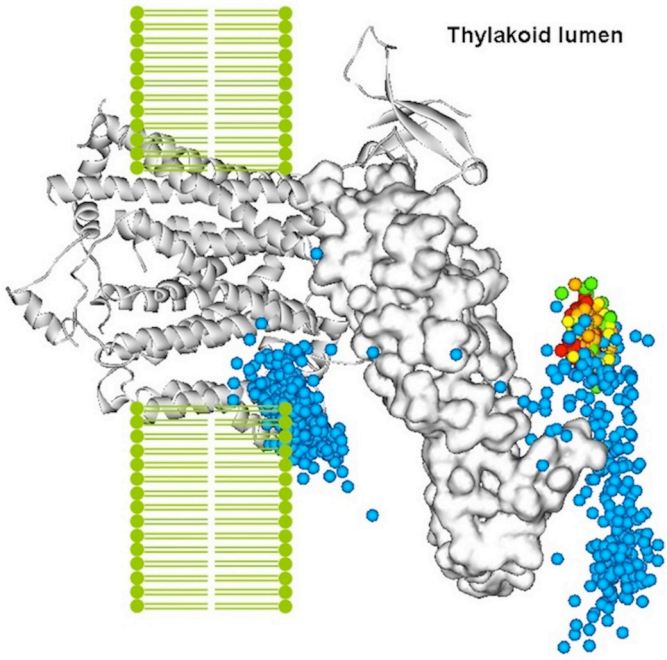

Figure 3.13. Encounter complex of the Pc-Cyt *f* complex superimposed on the structure of the b_6f complex (PDB entry 2ZT9), embedded in the thylakoid membrane. b_6f complex is a dimer, but only a monomer is shown. Cyt *f* is shown as a white surface and Pc centers-of-mass are represented by spheres, which are color-coded to indicate the distance between Cu in Pc and Fe in Cyt *f*, like in Figure 3.12B. The remaining components of the b_6f complex are shown as ribbons.

The encounter complex located opposite to the ET site places Pc inside the thylakoid membrane. The Cyt *f* soluble part was shown to be only lightly flexible in the thylakoid lumen,[164] suggesting that Pc cannot interact with Cyt *f* at this site under the physiological conditions. The observation of encounters at this site is likely an artifact due to the use of the soluble part of Cyt f in isolation and not embedded in the thylakoid membrane. The encounters located in front of the binding site diffusely extend from the haem to the end of the small domain of Cyt *f*. In other studies, diffusive encounter complexes have also been described.[165] Either the partners sample large areas or form encounters at several more defined regions on the surface of the partner.[24,30,136,166] Due to the diffuse distribution of negative charges on Cyt *f*, Pc probes a large part of the surface, instead of being attracted to specific sites by localized charges.

The formation of an encounter complex in many cases does not lead to a productive association, hence fruitful and futile encounter complexes can be distinguished.[24] The encounter ensemble shows a high density in the region directly in front to the haem, suitable for ET (red dots in Figure 3.12B). These orientations can obviously be considered fruitful encounters. For many other encounters, the distances between the redox centers are unfavorable for ET, but not all of these are futile. The further away from the ET active site Pc binds the less chance it has to diffuse to an ET active orientation before dissociation from Cyt *f*. Thus, the distinction between fruitful and futile encounter is a gradual one and it is difficult to tell when these complexes can rearrange into a productive orientation.

The population of the encounter state, in relation with the final state population, significantly varies among different complexes. In some non-ET complexes the population of the encounter state has been determined to be 10%,[27,140] in the ET complex Cyt *c*- Cyt *c* peroxidase it has been shown to be 30%.[28] For Pc-Cyt *f* complex from *Nostoc* it was not possible to establish the fraction accurately. Complexes, which only exist in the encounter state have also been described for myoglobin-cytochrome b_5[72] and adrenodoxin-cytochrome *c*.[30] In *Nostoc*, the existence of the complex as merely an encounter state seems unrealistic. The size of CSP in the wild type complex[98] and in the presence of spin labeled Cyt *f* also in this study is indicative of the formation of a stereo-specific complex. Moreover, PCS from the haem generated a converged structure stabilized in a defined orientation,[98] in which hydrophobic contacts and electrostatic interactions are optimized within the structure. The sensitivity of this approach to lowly populated states is limited, but it clearly demonstrates the existence of a final state.

At the same time, the diffuse nature of the encounter complex on Cyt *f* surface suggests that in this system a final orientation may not be a fundamental requirement for the functionality of the complex. In fact, the efficient turn-over required for rapid ET through the photosynthetic redox chain[167] precludes the formation of a tight complex and favors the conditions for the existence of the ET active complex in multiple orientations[4] to enhance the probability of ET.[21] Against this background the finding that hydrophobic contacts play a role not only in the final complex, but also in the encounter complex is interesting. It blurs the distinction between both states and would allow for a smooth transition from encounter to final complex via a gradual optimization of the hydrophobic contacts in the interface (Figure 3.14, solid line). This model of protein complex

formation allows for more rapid formation of the final complex than in the case of a model with an activation energy burrier between both states (Figure 3.14, dashed line), used for other protein complexes.[24]

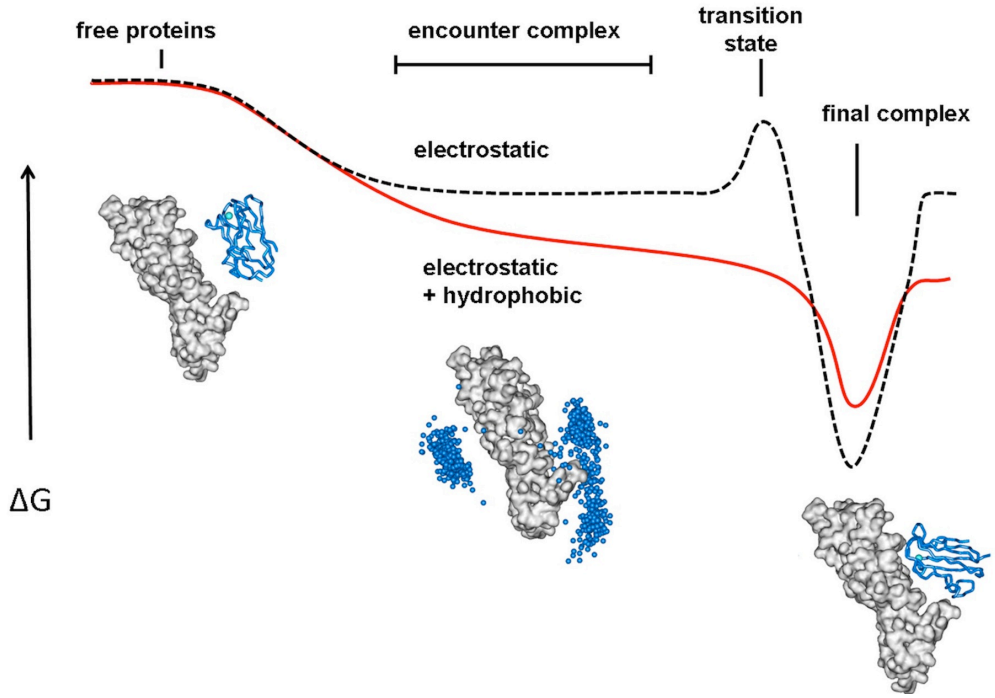

Figure 3.14. Free energy diagrams of proteins association pathways with high (dashed black line) and no (solid red line) energy barrier for the transition state from the encounter to the final complex.

In conclusion, we speculate that the hydrophobic interactions in the encounter complex may ensure a relative flat energy landscape during all phases of association, without a clear distinction between the encounter and the active complex. A flat energy landscape ensures rapid transitions between all states, which is relevant for systems that do not require a high degree of specificity, such as ET complexes.

Chapter IV

Loss of electrostatic interactions causes increase of dynamics within the plastocyanin-cytochrome *f* complex

Adapted with permission from
Scanu S., Foerster J.M., Timmer M., Ullmann G.M. and Ubbink M. (2013) Loss of electrostatic interactions causes increase of dynamics within the plastocyanin-cytochrome *f* complex. *Biochemistry,* August 28. Copyright 2013 American Chemical Society

Abstract

The studies presented in Chapter III on the photosynthetic electron transfer complex formed by cytochrome *f* and plastocyanin from *Nostoc sp.* PCC 7119 revealed that both hydrophobic interactions and electrostatic interactions, play a role in the process of complex formation. To study the balance between these two types of interactions in the encounter state and the final complex, the interaction between plastocyanin from *Phormidium laminosum* and cytochrome *f* from *Nostoc* was investigated, using NMR spectroscopy and Monte Carlo docking computations. This plastocyanin is a natural variant of that from *Nostoc*, in which the net positive charge of the protein is reverted to negative. Cytochrome f has a highly negative charge, and thus it was expected that the electrostatic interactions become unfavorable. NMR titrations indicated that a complex with an affinity intermediate between those of the *Nostoc* and *Phormidium laminosum* complexes is still formed. The orientation of plastocyanin in the complex was determined using pseudocontact shifts, demonstrating that the hydrophobic patch is the main site of interaction on plastocyanin similar to the head-on orientation found for the *Phormidium* complex. However, the interaction in the cross complex is dependent on electrostatic interactions, contrary to the *Phormidium* complex. The negative charge of plastocyanin decreases, but not abolishes, the attraction to cytochrome *f*, resulting in the formation of a more diffuse encounter complex than *Nostoc*, as could be determined using paramagnetic relaxation spectroscopy. This work illustrates the subtle interplay of electrostatic and hydrophobic interactions in complex formation of transient protein complexes. The results are discussed in the context of a model for complex formation on the basis of hydrophobic contacts in the encounter state.

Introduction

Protein association involves the formation of a dynamic encounter complex that is in equilibrium with the final, single-orientation complex.[1] In the encounter state, the proteins sample the surface of the partner, thus reducing the dimensionality of the search for the specific binding site.[21] Protein complex formation has been commonly described with a general model, in which the formation of the encounter complex is dominated by long-range electrostatic interactions, whereas the final state is determined by short-range interactions.[22] However, theoretical staudies demonstrated that desolvation can be a dominant interaction in the process of complex formation for systems with weak charge complementarity.[139,167,170] Furthermore, partial desolvation of the binding interface was reported for some encounter complexes.[140,171] The recent characterization of the encounter state of cytochrome *f* (^{N}Cyt *f*) and plastocyanin (^{N}Pc) complex from the cyanobacteria *Nostoc* sp. PCC 7119 ($^{N\text{-}N}$complex) demonstrated experimentally that electrostatic interactions alone cannot describe the encounter complex, suggesting that hydrophobic interactions also contribute to its formation (Chapter III). In the proposed model, long-range electrostatics result in the preorientation of ^{N}Pc relative to ^{N}Cyt *f* and hydrophobic interactions stabilize the encounter complex by promoting the overlap of the extended non-polar surfaces of both proteins. ^{N}Pc can diffuse in the hydrophobic interface and smoothly reach orientations capable of ET. The identification of hydrophobic interactions in the encounter state contrasts the view in which short-range interactions occur only in the final complex.[21]

Pc and Cyt *f* are photosynthetic redox partners in oxygenic organisms, such as plants, green algae and cyanobacteria. Pc is a soluble electron carrier, which shuttles electrons from Cyt *f* of the cytochrome b_6f complex to photosystem I.[4,167,168] The association of Pc and Cyt *f* is on the border between electrostatic-assisted[23] and desolvation-mediated association,[169] therefore representing a good model to elucidate the balance between electrostatic and hydrophobic interactions in protein complex formation. *In vitro*, electrostatic interactions enable fast association[7,8,59,60,66,69] and non-polar interactions favour the stabilization of the complex to in an ET active conformation.[46,54,98] The characterization of Pc-Cyt *f* complexes from several organisms revealed that small differences in the electrostatic surface properties of the individual proteins strongly influence both the binding equilibrium and the final orientations of the complexes. Both in plants[46,62,172] and in the cyanobacteria *Nostoc*[98] and *Prochlorothrix hollandica*,[32] electrostatic interactions influence the final orientation of Pc with respect to Cyt *f* within the complex and tilt the long side of Pc towards the small domain of Cyt *f* in the so-called side-on orientation. The complex from the cyanobacterium *Phormidium laminosum* ($^{Ph\text{-}Ph}$complex) was found instead in the head-on orientation, in which solely the hydrophobic ET site represents the binding site.[54] Neutralization of charged residues on the surface of ^{Ph}Pc[7] and ^{N}Pc[8] has shown to have greater effect on the kinetics of the reaction than similar modifications on ^{Ph}Cyt *f*[69] and ^{N}Cyt *f*,[66] respectively. ^{Ph}Cyt *f* and ^{N}Cyt *f* are electrostatically similar, with an overall charge of -13 and -15,[49] respectively, and a rather even distribution of the negative charges over the surfaces. The two Pcs

show 63% of amino acid sequence identity and very similar three-dimensional structures, but they vary considerably in their electrostatic properties. ^{N}Pc is overall positively charged with six lysines (K6, K11, K20, K24, K35 and K100) forming an extended charged patch, which juxtaposes the long side of ^{N}Cyt *f* in the side-on orientation. In ^{Ph}Pc, K11 and K20 are substituted by serine and asparagine, respectively, and the positively charged patch is composed by four lysines (K6, K30, K35 and K100), yielding a protein with a net negative charge (-1 at pH 6). To evaluate the effects that these electrostatic differences between the two Pcs can cause along the association pathway of Cyt *f* and Pc, the complex of ^{N}Cyt *f* and Zn-substituted ^{Ph}Pc ($^{N\text{-}Ph}$complex) was studied using NMR spectroscopy and computational approaches. The consequences of this on binding affinity, final complex orientation and encounter complex will be discussed in the light of the findings for the $^{N\text{-}N}$complex presented in Chapter III.

Experimental section

Protein production and purification

The plasmid pET11PC,[172] which contains the gene for wild type ^{Ph}Pc, was transformed in *E.coli* BL21 pLysS. ^{15}N enriched-Zn substituted Pc was produced as described before for ^{N}Pc,[29] with the difference that ampicilline (100 mg/L) and chloramphenicol (20 mg/L) were added to the growth media in stead of kanamycin (25 mg/L). The purification procedure was reported before.[54] The concentration of the protein was determined by absorbance spectroscopy using ε_{280} = 5 mM^{-1}cm^{-1}. The yield of pure protein was 4 mg/L of culture.
The pEAF-WT plasmid, containing the gene of the soluble domain (residue 1-254) of *Nostoc* sp. PCC7119 Cyf *f* was kindly provided by Prof. Dr. Miguel A. De la Rosa (University of Seville). Cyt *f* mutants were obtained using pEAF-WT plasmid as template for mutagenesis as described before (Chapter III).[29,143] Production and purification of the protein, and spin label attachment were performed as previously reported (Chapter III).[8,29]

NMR experiments

All NMR samples contained MES (20 mM, pH 6) and 6% D_2O for lock. The pH of the sample was adjusted with small aliquots of HCl (0.5 M) and NaOH (0.5 M). For the chemical shift perturbation (CSP) experiments Cyt *f* was titrated into Zn-substituted ^{15}N Pc (40 μM). Spectra were recorded at multiple Cyt *f*:Pc molar ratios (0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 2.5, 5.0, 7.5, 10). For measurements of the PCSs, HSQC spectra of the free Pc and in the presence of ferric and ferrous Cyt *f* were acquired on the same sample. Ferric Cyt *f* was oxidized with $K_3[Fe(CN)_6]$ and loaded onto a PD10 column to remove the oxidant, concentrated and then added to Pc (final concentration 135μ M) to Cyt *f*:Pc molar ratio 3:1. Ferric Cyt *f* was then reduced by adding 10 molar equivalents of ascorbic acid directly into the sample. For the PRE experiments the ferric state of Cyt *f* was preserved by addition of $K_3[Fe(CN)_6]$

(50 μM). These samples contained 135 μM Cyt *f* for Q125C mutant and 300 μM for the other mutants, labelled with either MTS or MTSL. Samples also contained Zn-substituted ^{15}N Pc, 45 μM in the complex with Q125C Cyt *f* and 100 μM for the other Cyt *f* mutants. All NMR spectra were recorded at 298 K on a Bruker Avance III 600 MHz spectrometer equipped with a TCI-Z-GRAD CryoProbe. The ^{1}H-^{15}N HSQC spectra were acquired with 1024 and 80 complex points in the direct and indirect dimensions, respectively.

NMR data analysis

The NMR spectra were processed with NmrPipe[145] and analyzed with CcpNMR Analysis.[146] CSP analysis was carried out as described before.[29] PCS was defined as the chemical shift difference for a resonance in the presence of paramagnetic and diamagnetic Cyt *f*, according to previously reported procedures.[46,98]
The PREs were determined according to the procedure described by Battiste and Wagner.[81] The intensity ratio I_p/I_d of the Pc resonances in the presence of MTSL-Cyt *f* (I_p) and MTS-Cyt *f* (I_d) were normalized by dividing them by the average value of the ten largest I_p/I_d values (1.09 for Q7C, 1.05 for Q38C, 2.21 for N71C, 1.41 for Q125C, 1.16 for S181C and 1.25 for S192). The PRE (Γ_2) values were calculated according to Equation 4.1:

$$\frac{I_p}{I_d} = \frac{R_{2d}\exp\left(-\Gamma_2 t\right)}{R_{2d}+\Gamma_2} \qquad (4.1)$$

R_{2d} represents the transverse relaxation rate in the diamagnetic sample, which was calculated from the linewidth at half height obtained from a Lorentzian peak fit in the direct dimension, by using FUDA (this software was kindly provided by Dr. D. Fleming Hansen, University College London). The symbol *t* indicates the time for transverse relaxation during the pulse sequence (9 ms). The Γ_2 values were extrapolated to the 100% bound state using the experimentally obtained K_D. The uncertainty for I_p/I_d ratios ($\Delta\sigma_{Ip/Id}$) was determined by error propagation according to Equation 4.2:

$$\Delta\sigma_{\frac{I_p}{I_d}} = \frac{I_p}{I_d}\sqrt{\left(\frac{\sigma_p}{I_p}\right)^2+\left(\frac{\sigma_d}{I_d}\right)^2} \qquad (4.2)$$

In which σ_p and σ_d represent the noise level of paramagnetic and diamagnetic spectra, respectively. The noise level of each spectrum is represented by the standard deviation of the intensities measured at ten randomly chosen positions between the resonances.

Docking calculations

The structure of the soluble part of Cyt *f* (residues 1-254) used for the calculation was taken from PDB entry 2ZT9[147] as described before.[29] The structure of ^{Ph}Pc was

taken from PDB entry 2Q5B. The orientation of ^{Ph}Pc in complex with ^{N}Cyt *f* was determined by rigid body docking using solely PCS restraints with the option PARAstraints[174] in Xplor-NIH 2.9.9.[175] For this reason the observed $^{1}H\Delta\delta_{PCS}$ were extrapolated to 100% bound Pc by dividing them by the fraction bound (0.47). The size of the axial magnetic component of the magnetic susceptibility anisotropy tensor ($\Delta\chi_{ax}$), was derived from the g-tensor values measured by EPR spectroscopy on plant Cyt *f* (7×10^{-32} m^3).[98] However, the value required to obtain convergence of the structure calculations is much smaller. One reason for this is the temperature difference between the EPR measurements (10 K) and NMR spectra (taken at 298 K). The second reason is averaging effect occurring in the encounter state, which reduces the PCS considerably. The $\Delta\chi_{ax}$ was varied from 0.61 to 3.3×10^{-32} m^3. The best convergence was found for $\Delta\chi_{ax} = 0.87 \times 10^{-32}$ m^3, whereas it was taken to be 7×10^{-32} m^3 in the $^{N\text{-}N}$complex.[98] The intermolecular PCSs from the ferric haem iron of Cyt *f* to the backbone amide atoms in Pc were back calculated from the best 20 structures and compared with the experimental PCSs. Equation 4.3 was used for the PCSs calculation, assuming an axial magnetic susceptibility tensor oriented along the vector defined by the iron and the N-atom of Y1 of Cyt *f*:[46]

$$\Delta\delta_{PCS} = \frac{\Delta\chi_{ax}}{12\pi r^3}\left(3\cos^2\theta - 1\right) \quad (4.3)$$

In which $\Delta\delta_{PCS}$ is the PCS, r is the distance between haem iron and observed Pc nucleus, and θ is the angle between Pc nucleus, haem iron and the nitrogen of the amine group of Y1 in Cyt *f*. The $\Delta\chi_{ax}$ was also varied in the backcalculation of the PCSs, and the best fit between the average back calculated PCSs from the best 20 structures representing the final complex model and the experimental data was found for $\Delta\chi_{ax} = 0.87 \times 10^{-32}$ m^3 as well. The degree of agreement between observed (PCS^{obs}) and calculated (PCS^{calc}) PCSs was determined by the PCS Q factor, defined as:

$$Q_{PCS} = \sqrt{\frac{\Sigma\left(PCS^{obs} - PCS^{calc}\right)^2}{\Sigma\left(\left|PCS^{obs}\right| + \left|PCS^{calc}\right|\right)^2}} \quad (4.4)$$

The ensemble docking was performed as described for the $^{N\text{-}N}$complex (Chapter III). The restraints for the calculations were obtained according to Equation 4.5:

$$\Gamma_2^{obs} = f_1\Gamma_2^{ens} + f_2\Gamma_2^{final} \qquad f_1 + f_2 = 1 \quad (4.5)$$

The ensemble Γ_2 (Γ_2^{ens}) was calculated as the difference between observed Γ_2 (Γ_2^{obs}) and average back calculated Γ_2 from the model of the PCS-based final complex models (Γ_2^{final}). The calculations were carried out with f_2 values of = 0, 0.15, 0.25, 0.35, 0.5, 0.65, 0.75, 0.85, 0.95 and 1.0. The restraints were grouped into three classes as described before.[29] For the visualization of the encounter

complex ensemble 150 docking were performed, yielding 148 ensembles of seven Pc conformers, with a difference in the total restraint energy ≤20%.
The ensembles from separated dockings were evaluated by means of the averaged violation for all experimental restraints as described before (Chapter III). The intermolecular PCSs from the ferric haem iron of Cyt *f* to the backbone amide atoms of all Pc conformers (PCS^{ens}) were back calculated with Xplor NIH 2.9.9. PCS^{ens} were linearly combined with the back calculated PCSs from the final model (PCS^{final}) using different population fraction of the encounter complex (f_1) according to Equation 4.6:

$$PCS^{calc} = (1 - f_1)\, PCS^{single} + f_1\, PCS^{ens} \qquad (4.6)$$

The agreement between observed (PCS^{obs}) and calculated (PCS^{calc}) PCS was given by the PCS Q factor, as defined in equation 4.4.

Monte Carlo simulations of the encounter complex

The structure files for Cyt *f* and Pc were taken from the PDB entries 2ZT9[147] and 2Q5B, respectively. Monte Carlo simulations generate a Boltzmann distribution of encounter complexes according to their electrostatic-interaction energy.[57] The simulations were performed using a previously described approach (Chapter III).[28]

Results and discussion

Affinity and binding site

For the characterization of the non-physiological cyanobacterial $^{N\text{-}Ph}$complex formed by ^{N}Cyt *f* and ^{Ph}Pc, ^{15}N enriched Zn ^{Ph}Pc was titrated to either oxidized ^{N}Cyt *f* (Fe^{III}) or reduced ^{N}Cyt *f* (Fe^{II}) to molar ratios Pc:Cyt *f* 1:10 and 1:3, respectively. ^{Ph}Pc was produced with Zn rather than the Cu in the binding site to avoid the interference of ET and the disappearance of important resonances due to the line-broadening caused by the paramagnetic Cu.[155] Each titration point was monitored through the acquisition of ^{15}N-^{1}H HSQC spectra. Upon addition of Cyt *f* a number of resonances shifted in the spectrum, indicating complex formation. The appearance of shifting resonances indicates that free and bound Pc are in fast exchange on the NMR time scale. The binding curves for the most affected residues were obtained by plotting the CSP ($\Delta\delta_H$) versus Cyt *f*(Fe^{III})/Pc molar ratio, as shown in Figure 4.1A.

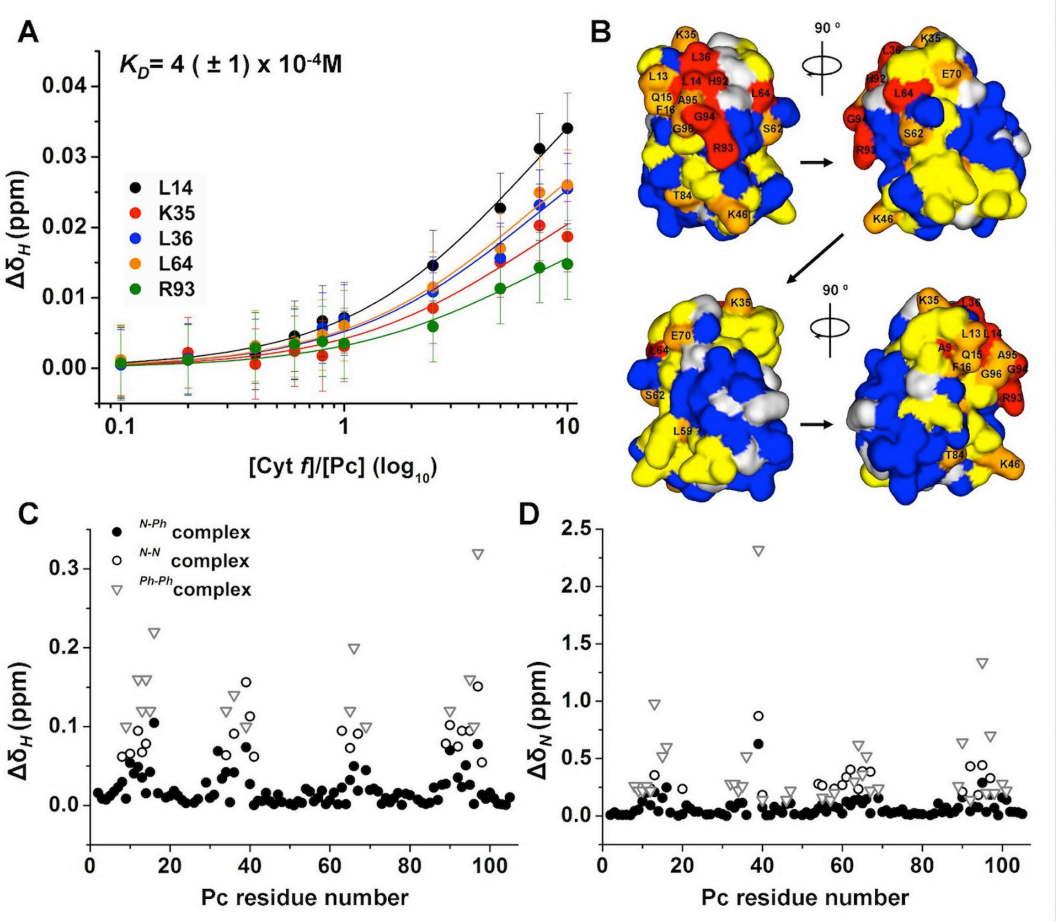

Figure 4.1. The interaction of Zn substituted ^{15}N ^{Ph}Pc with ^{N}Cyt *f*. A) CSP curves for Zn ^{Ph}Pc binding to ^{N}Cyt *f* for selected residues fitted to a 1:1 interaction model. B) Binding map of ^{Ph}Pc in the presence of wild type ^{N}Cyt *f* (Fe III), color-coded on a surface model of Pc (PDB entry 2Q5B). The red color corresponds to $\Delta\delta_{AVG} \geq$ 0.030 ppm, orange to $\Delta\delta_{AVG} \geq$ 0.015 ppm, yellow to $\Delta\delta_{AVG} \geq$ 0.0075 ppm, blue to $\Delta\delta_{AVG} \leq$ 0.0075 ppm. Prolines and overlapping residues are colored in light grey. (C, D) CSPs of ^{Ph}Pc resonances upon binding of ^{N}Cyt *f* (Fe II). The CSPs in ^{1}H dimension (C) and in ^{15}N dimension (D) observed for the $^{N\text{-}Ph}$complex are shown as black dots, for the $^{N\text{-}N}$complex as black circles and for $^{Ph\text{-}Ph}$complex as grey triangles.

The CSP curves did not reach saturation, indicating a low affinity. The global fit of the binding curves to a 1:1 binding model yielded a dissociation constant of 4 (± 1) × 10^{-4} M. This value is in between the reported values for the $^{Ph\text{-}Ph}$complex and $^{N\text{-}N}$complex, being ≈10 × 10^{-4} M[54] and 0.8 × 10^{-4} M,[29] respectively. Whereas the cross complex formed by ^{Ph}Cyt *f* and ^{N}Pc ($^{Ph\text{-}N}$complex) was reported to have similar affinity to $^{N\text{-}N}$complex (K_D= 0.8 × 10^{-4} M),[49] the $^{N\text{-}Ph}$complex shows an affinity intermediate to that of the two physiological complexes but closer to $^{Ph\text{-}Ph}$complex. The experimental K_D was used to determine that the fraction of ^{Ph}Pc bound to ^{N}Cyt *f* at the last point of the titration was 0.52 and the $\Delta\delta_{AVG}$ were extrapolated to the 100% bound form. The CSP map of Zn ^{Ph}Pc was obtained by color-coding each residue according to the size of $\Delta\delta_{AVG}$ (Figure 4.1B). The largest

effects were observed for residues surrounding the metal binding site, namely A9, L14, L36, H39, L64, H92, R93 and G94, colored in red. Most of these residues are hydrophobic and make up the hydrophobic patch of Pc, which was also identified as main binding site in the structural models of the $^{N\text{-}N}$complex and $^{Ph\text{-}Ph}$complex.[54,98] Clearly, the hydrophobic patch plays a fundamental role in the formation of the Cyt *f*-Pc complexes. The CSP map is qualitatively similar to that of the $^{N\text{-}N}$complex[29,98] with a prominent perturbation for R93, known to be involved in the binding in both the $^{N\text{-}N}$complex[8,66,98] and $^{Ph\text{-}Ph}$complex.[7] Interestingly, a significant CSP was also observed for K46 in the $^{N\text{-}Ph}$complex. K46 is located far from the hydrophobic patch, well below R93, and kinetic studies suggested its implication in the electrostatic modulation of the binding of $^{Ph\text{-}Ph}$complex.[67]

Structure of the final complex

The orientation of ^{Ph}Pc in complex with ^{N}Cyt *f* was determined by taking advantage of the intermolecular PCSs caused by the paramagnetic oxidized iron of Cyt *f* on Pc backbone amide protons, in a similar way as was done previously for other Pc-Cyt *f* complexes.[32,46,54,62] PCSs arise from the through-space dipolar interaction between the spin of the unpaired electron and that of the observed nucleus. PCS is distance and orientation dependent and provides restraints for structural calculations. The calculations converged to an ensemble of structures. The best 20 structures exhibit a difference in the restraint energy of less than 6% and are shown in Figure 4.2A.

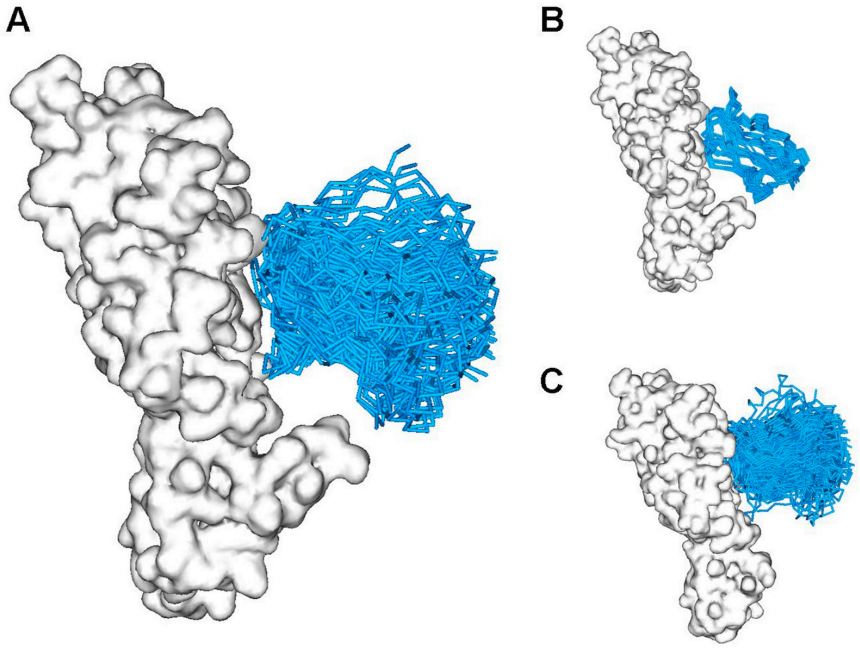

Figure 4.2. Comparison of the structures of Pc-Cyt *f* complexes, showing the structure obtained for the $^{N\text{-}Ph}$complex (A), and the physiological $^{N\text{-}N}$complex (PDB entry 1TU2[98]) (B) and $^{Ph\text{-}Ph}$complex (C). Cyt *f* is shown as a white surface model on the PDB entry 2ZT9 and Pc is represented by the ensemble of the 20 (A), 10 (B) and 25 (C) lowest energy conformations, shown as cyan C_α traces.

The resulting model shows a high degree of variability, but in all structures the hydrophobic patch of Pc is making contact with the hydrophobic patch surrounding the haem of Cyt *f* and represents the entire complex interface. All structures showed an interaction between H92 of Pc and F3 of Cyt *f*, also found in the *Ph-Ph*complex[54] and *N-N*complex.[98] The binding interface is composed of polar and hydrophobic residues, located in the regions 11-14, 36-39, 64-68 and 90-95 on *Ph*Pc. R93 represent the only charged interfacial residue. The averaged Cu-Fe distance in the ensemble was 15.3 (± 0.5) Å. In Figure 4.3A, the observed (black filled circles) and the back calculated PCSs for the best 20 structures (grey lines) are plotted versus Pc residue numbers.

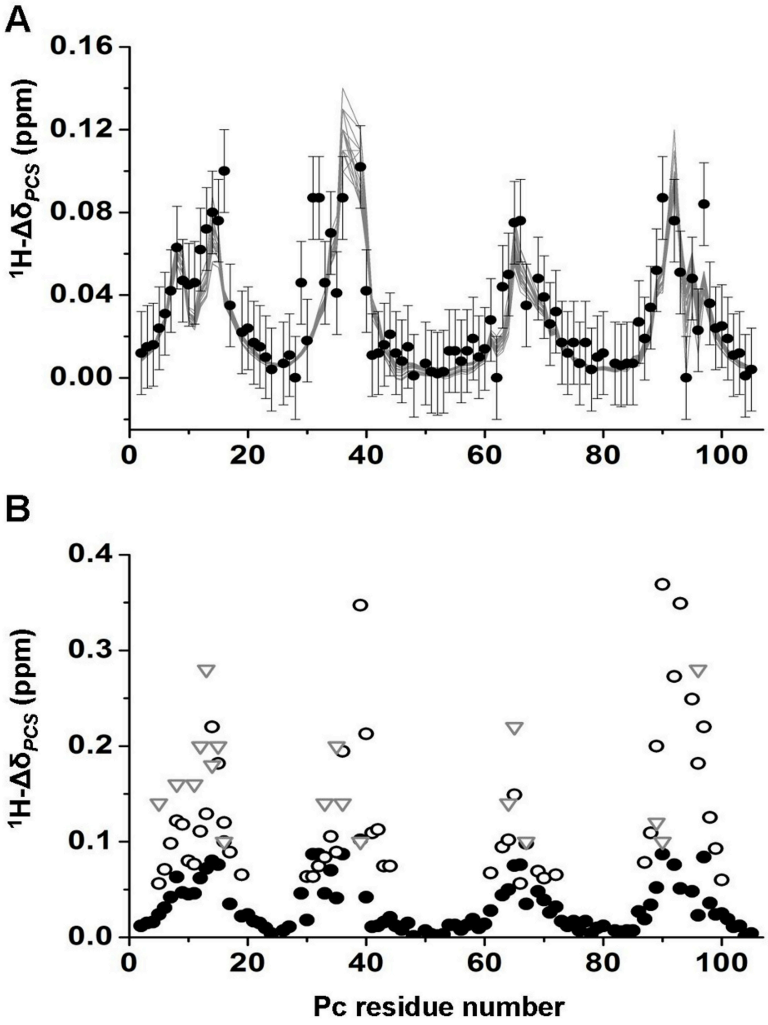

Figure 4.3. Evaluation of *N-Ph*complex. In panel A, the observed ^{1}HΔδ$_{PCS}$, which were extrapolated to 100% bound Pc, are shown as black filled circles and the back calculated ^{1}HΔδ$_{PCS}$ for the 20 lowest PCS energy structures are shown as grey lines. The error bars represent the estimated experimental errors in the resonance positions. In panel B, the observed ^{1}HΔδ$_{PCS}$ for the *N-Ph*complex are shown as black filled circles, for the *N-N*complex[98] as black open circles, and for the *Ph-Ph*complex[54] as grey triangles. All PCSs were extrapolated to the 100% bound form and plotted versus Pc residue numbers.

For most residues experimental and back calculated PCSs agree within the error margins. Small deviations are observed for F16, V29, W31, V32 and A90, which form the edge of the hydrophobic binding site, and M97, which coordinates the metal. Considering the relative vicinity of these residues to the haem, it is possible that the approximations made for the size, axiality and orientation of the magnetic susceptibility tensor cause these deviations. The overall quality of the structures was evaluated by calculating a quality (Q) factor for the back calculated PCSs for each structure of the final model and the experimental PCSs (See Experimental Section, Equation 4.4). The averaged Q value was calculated to be 0.23 (± 0.1).

The orientation of Pc in the complex is most similar to the head-on orientation found in the $^{Ph\text{-}Ph}$complex (Figure 4.2C) rather that the side-on orientation of the $^{N\text{-}N}$complex (Figure 4.2B). ^{Ph}Pc is oriented perpendicular to the haem with the hydrophobic patch at a slight angle to the small domain of ^{N}Cyt *f*. In the $^{N\text{-}N}$complex the specific electrostatic contacts between K57 and K62 of ^{N}Pc and E189 and D64 of ^{N}Cyt *f* appears to be responsible for the long side of Pc to be tilted toward Cyt *f*.[49,98] These lysines are substituted with D57 and S62, respectively, in ^{Ph}Pc and the loss of these important electrostatic contacts may lead to the "head-on" orientation in the $^{N\text{-}Ph}$complex. In the $^{N\text{-}Ph}$complex, only the bottom part of ^{Ph}Pc (relative to the hydrophobic patch) is turned toward the small domain of ^{N}Cyt *f*, probably as a consequence of the charge-charge interaction between K46 of ^{Ph}Pc and E189 and D190 of ^{N}Cyt *f*. The soluble part of ^{Ph}Cyt *f* is shorter than ^{N}Cyt *f*, comprizing 249 instead of 254 residues. This causes the small domain to be less extended and not in direct contact with Pc in the $^{Ph\text{-}Ph}$complex.[54] In the $^{N\text{-}Ph}$complex K46 is in a favorable position to have electrostatic interactions with E189 and D190 in the prominent small domain of Cyt *f*.

Since PCSs depend on the orientation of the observed nucleus with respect to the paramagnet, the presence of multiple orientations is expected to influence the size of PCSs. In Pc-Cyt *f* complex from *Prochlorothrix hollandica,* the mutation of Y12 and P14 in Pc to Gly and Leu, respectively, caused an increase of dynamics, as judged by the decrease of PCSs for nuclei in certain regions of Pc.[32] In Figure 3B, the observed $^{1}H\Delta\delta_{PCS}$ of $^{N\text{-}Ph}$complex were compared with the reported values for $^{N\text{-}N}$complex[98] and $^{Ph\text{-}Ph}$complex,[54] each extrapolated to the 100% bound state. The pattern of the $^{1}H\Delta\delta_{PCS}$ is similar for all complexes, but the sizes of $^{1}H\Delta\delta_{PCS}$ are comparable only for the two physiological complexes, whereas they are considerably lower for the $^{N\text{-}Ph}$complex. This indicates that in the $^{N\text{-}Ph}$complex the dynamics of Pc is larger than in both the $^{N\text{-}N}$complex and the $^{Ph\text{-}Ph}$complex.

The encounter complex

To map the distribution of the encounter intermediates on ^{N}Cyt *f* surface in the $^{N\text{-}Ph}$complex, six spin labels were attached on ^{N}Cyt *f*, one at a time, and PREs were measured on the amide backbone protons of ^{Ph}Pc. Cyt *f* was added to Pc in a molar ratio Pc:Cyt *f* of 1:3. PRE causes line broadening of Pc resonances resulting in a low ratio of peak intensities in the spectra of the paramagnetic and diamagnetic samples (I_p/I_d). In Figure 4.4 (*central panel*) the positions of spin labels are shown on a surface model of Cyt *f* with respect to Pc oriented as found

in the lowest energy structure of the PCS-based final complex (cyan C_α trace). Spin labels attached to Cyt *f* on the same side as the binding site for Pc, at positions Q7, N71, and S192, caused a large decrease of I_p/I_d ratios of Pc resonances.

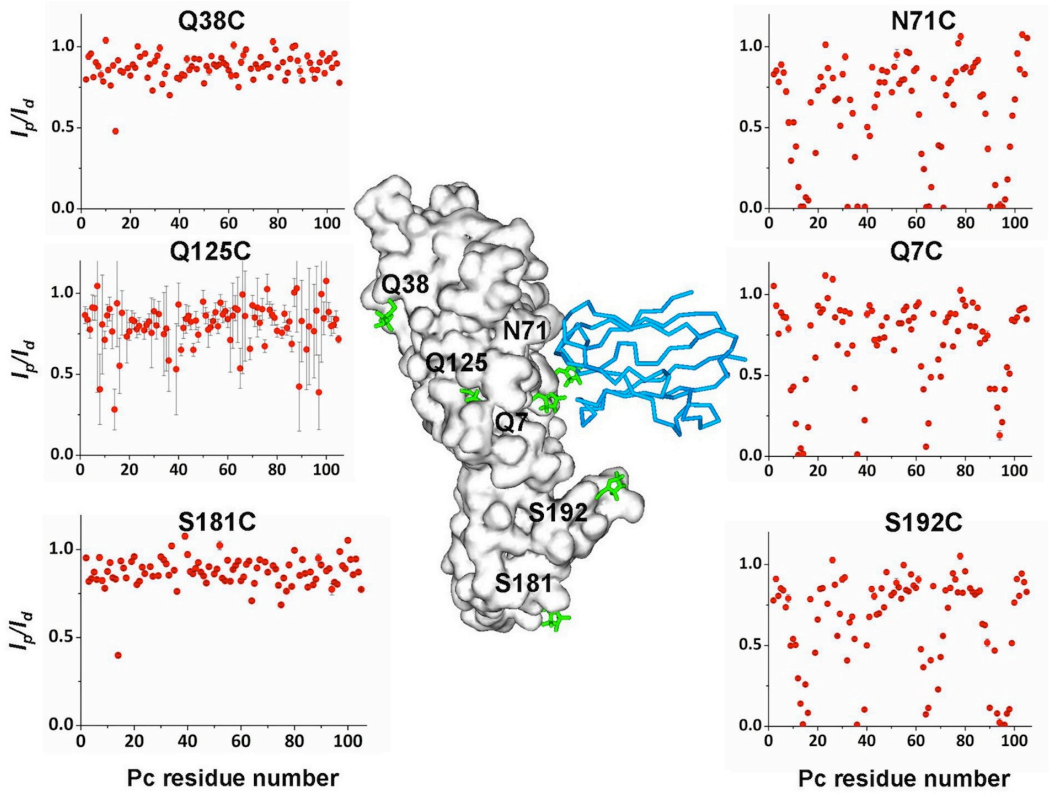

Figure 4.4. PRE in the $^{N\text{-}Ph}$complex. *Central panel*. Location of the spin labels (green sticks) modelled on the ^{N}Cyt *f* (PDB entry 2ZT9). Cyt *f* is shown as grey surface and Pc is represented as cyan C_α trace, oriented as the PCS-based final complex. *Side panels*. The I_p/I_d ratios are plotted versus the Pc residue number for each of the spin label positions on Cyt *f*. The error bars represent the uncertainty for I_p/I_d ratios based on the noise levels of the spectra. For most points, the error bar is within the symbol.

It is noteworthy that also spin labels attached on the backside of Cyt *f*, at the positions Q38 and Q125, or located far away, S181, showed a moderate to large decrease for some resonances. The large error bars calculated for the ratios in the presence of Q125C mutant are due to the lower concentration of Pc in this sample (45 μM) as compared to the other mutants (100 μM) resulting in a low signal-to-noise ratio. The I_p/I_d ratios were used to determine the PRE (Γ_2). In the fast exchange regime (see above) the observed PREs are weighted averages of free Pc, encounter complex and final complex. The PREs were extrapolated to the 100% bound state (encounter complex + final complex) by dividing by the fraction of bound Pc. The PREs caused by each spin label were mapped on the surface of Pc (Figure 4.5).

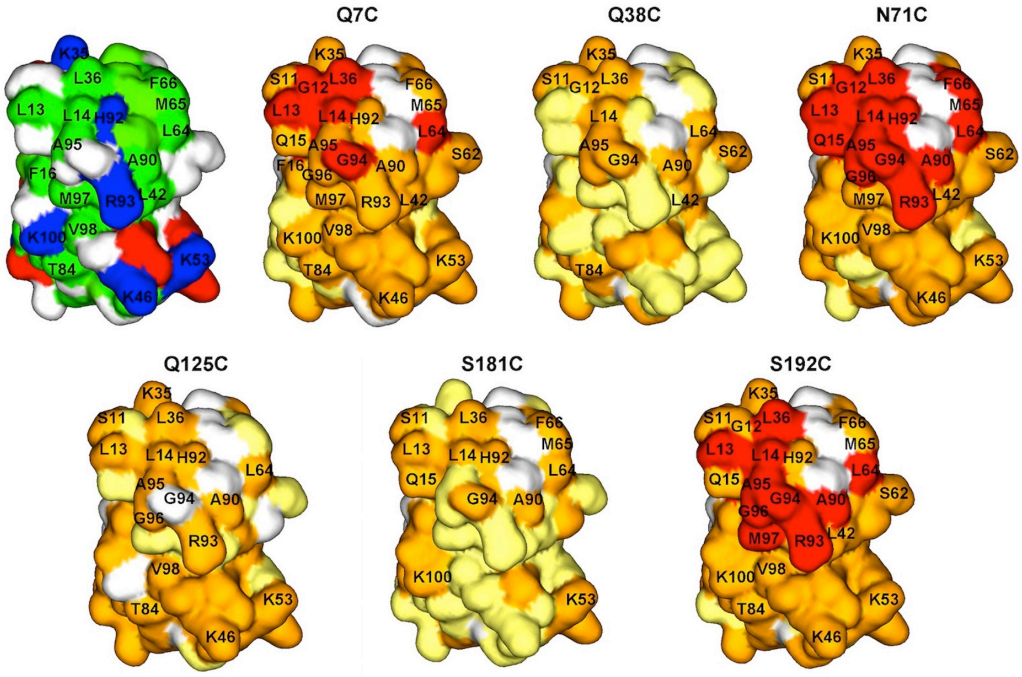

Figure 4.5. PRE maps of ^{15}N enriched-Zn substituted ^{Ph}Pc in the presence of MTSL-conjugated ^{N}Cyt *f*, color-coded on a surface model of Pc (PDB-entry 2Q5B). Experimental PREs were extrapolated to 100% bound Pc. Residues with $\Gamma_2 \geq 200$ s^{-1} are colored in red; with 10 s^{-1}< Γ_2< 200 s^{-1} in orange and with $\Gamma_2 \leq 10$ s^{-1} in light yellow. Prolines and residues with overlapping resonances are colored in white. On the top left, Pc is colored according to its charge distribution. Negatively and positively charged residues are shown in red and blue, respectively. Hydrophobic residues are in green and polar residues in white.

Even though the three spin labels located at the same side of Cyt *f* as the binding site (Q7C, N71C, S192C) are relatively far from each other, the PRE patterns are very similar and resemble the CSP map in the presence of wild type Cyt *f* (Figure 4.1B). The qualitative similarity of the PRE patterns suggests that Pc samples a large area of the Cyt *f* surface, while maintaining the same relative orientation to Cyt *f*. The highest PREs were observed for residues located in the hydrophobic patch of Pc, indicated as main binding site in the PCS-based final complex. Most of these residues are hydrophobic or polar, with the exception of R93 that was strongly affected by PRE in the presence of spin labels in N71 and S192. The same residue exhibited a high CSP in the presence of wild type Cyt *f* (Figure 4.1B). Interestingly, for most of these residues moderate PREs were also observed in the presence of spin labels attached to the backside of Cyt *f* with respect to the PCS-based binding site of Pc, indicating that Pc also visits this part of Cyt *f*.

The encounter complex was visualized by ensemble docking. This approach is based on the fact that PREs result from the weighted average contribution of all species in solution, which are either in the final or in the encounter orientations.[84] To represent all species that contribute to the observed PREs, multiple conformers of a protein are simultaneously docked on the other protein to obtain a population distribution that fits the experimental data. For the calculations the PREs are converted into distances and used as restraints for the docking. Each

docking yields a unique ensemble of orientations that account for the experimental PREs. The ensemble docking of ^{Ph}Pc and ^{N}Cyt *f* was essentially performed as described for the $^{N\text{-}N}$complex in Chapter III and the size of the ensemble was set to *N*=7, where *N* represents the number of copies of Pc in the ensemble. To separate the PRE contribution of the complex in the final state, the averaged back calculated PREs from the PCS-based models of the final complex were subtracted from experimental PREs and the resulting PREs were converted into distance restraints. A series of ensemble docking calculations was then carried out by varying the population of the final state (f_2) from 1-0. The resulting ensembles were evaluated by calculating the average distance violation over all experimental distances. The average distance violations were plotted versus the percentage of the encounter complex (Figure 4.6).

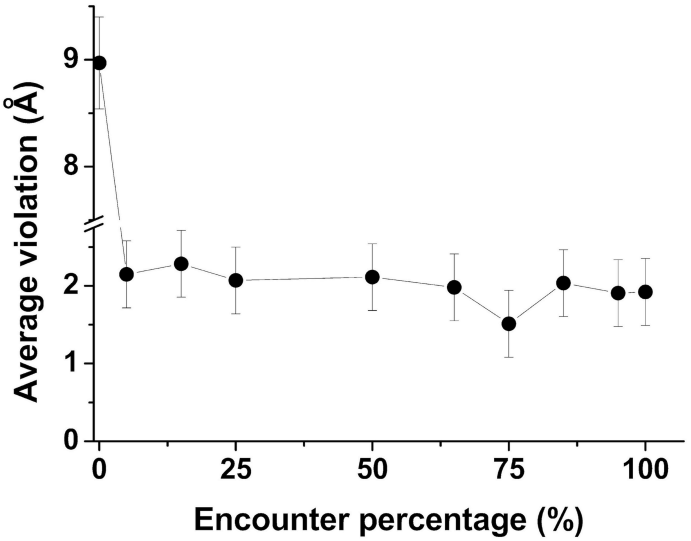

Figure 4.6. Plot of the average violation of all experimental distances versus the ensemble percentage included in the restraints for the calculations. Error bars represent 2 × SD of the average violations obtained from three independent calculations performed with N= 7 and f_2= 0.

The violations show that the observed PREs are not explained by the PCS-based structure alone. A significant decrease in the average violation is already observed when the encounter complex is taken to be 5%. Further increase of the encounter complex fraction in the restraints did not improve the fit of the data. For all generated ensembles an average violation of about 2 Å was observed.
The calculations for the representation of the encounter complex were performed assuming a pure encounter state (f_1=1). The comparison of the back calculated distances between the oxygen atom of the spin labels and the amide protons of all Pc conformers of the generated encounter complex (red line in Figure 4.7) and the back calculated distances in the PCS-based models of the final complex (blue line) shows that only the generated encounter complex fits the experimental PRE data (green dots and line).

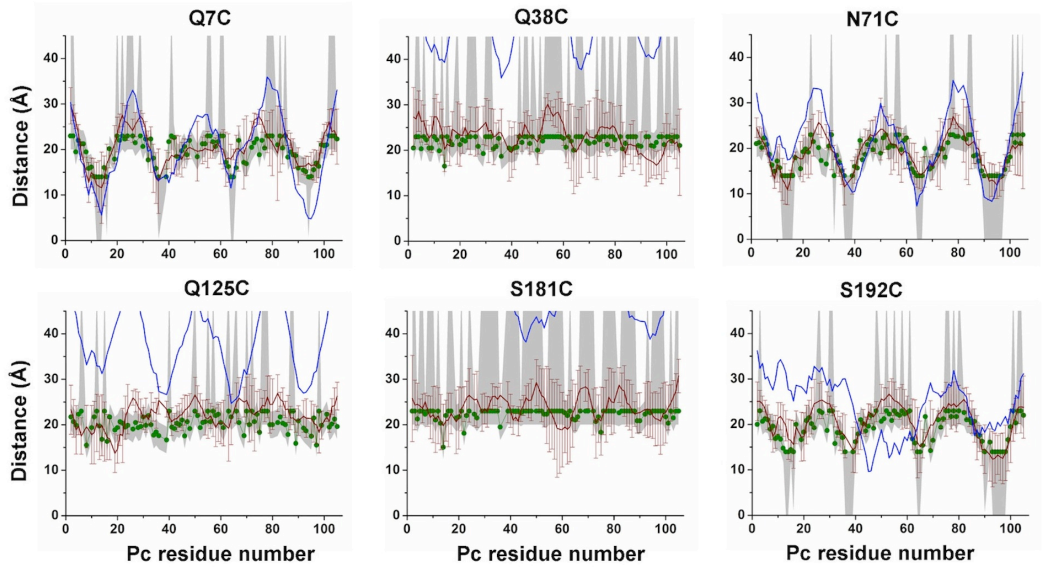

Figure 4.7. Ensemble docking. Experimental and back calculated average distances between Pc amide protons and oxygen atoms of MTSL conjugated to Cyt *f* are plotted against the Pc residue number. The green circles and lines represent the experimental distances and the grey areas indicate the error margins. The average distances back calculated from the 20 lowest-energy solutions of the PRE driven ensemble docking are shown as a red line with error bars representing the SD. The average back calculated distances from the PCS-based final complex models are shown as a blue line. Calculations were performed with $N = 7$ and $f_1 = 1$.

The main deviation is represented by S192, indicating that ^{Ph}Pc spends more time close to this spin label than expected from the PCS-based models. This suggests that PREs from S192 mainly arise from the encounter complex. Most of calculated distances from the generated encounter complex lie within the error margins of the experimental values. Deviations were observed for the spin label Q125C, likely due to the poor data quality (see above). Since in principle both PCS and PRE report on intra-complex dynamics, the PCSs from Cyt *f* iron to amide backbone protons of Pc were back calculated for the ensemble generated using PRE (PCS^{ens}). To determine the population fractions of the complex in the final and in the encounter states PCS^{ens} were linearly combined with the calculated PCSs (PCS^{final}) from the final models using different population fractions of the encounter complex (see Experimental Section, Equation 4.6). The different combinations were correlated to the experimental PCSs by means of a Q factor (Equation 4.4). To account for the uncertainty of the size of the axial component of the magnetic susceptibility anisotropy of Cyt *f* Fe^{III}, a range of $\Delta\chi_{ax}$ values was used for the determination of the Q values. In Figure 4.8, the Q factors are plotted versus the percentage of the encounter complex. As illustrated also in Chapter III, the fit the data depends on the size of $\Delta\chi_{ax}$ and a reliable estimation of the fraction of the encounter complex unfortunately cannot be made.

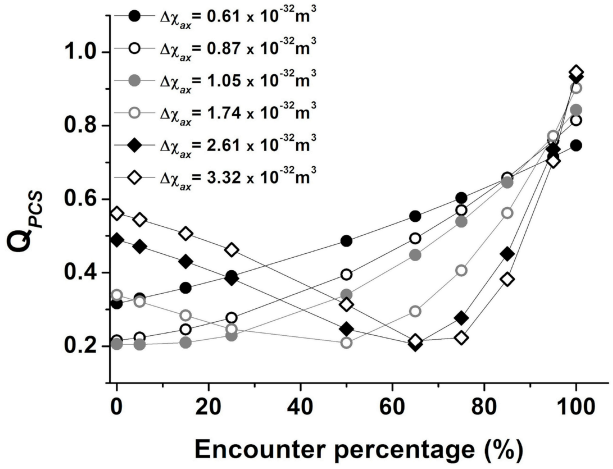

Figure 4.8. Q factors calculated for a combination of experimental PCSs measured for the specific complex and for different combination of back calculated PCSs from the final and the encounter complex. Q factors are plotted versus the percentage of the encounter complex. The Q factors were calculated at different values of the size of the axial component of the magnetic susceptibility tensor (see Experimental Section).

To represent the encounter complex, an ensemble from 145 docking solutions with *N*=7 and f_1=1, with a total of 1015 Pc conformers was generated (Figure 4.9A).

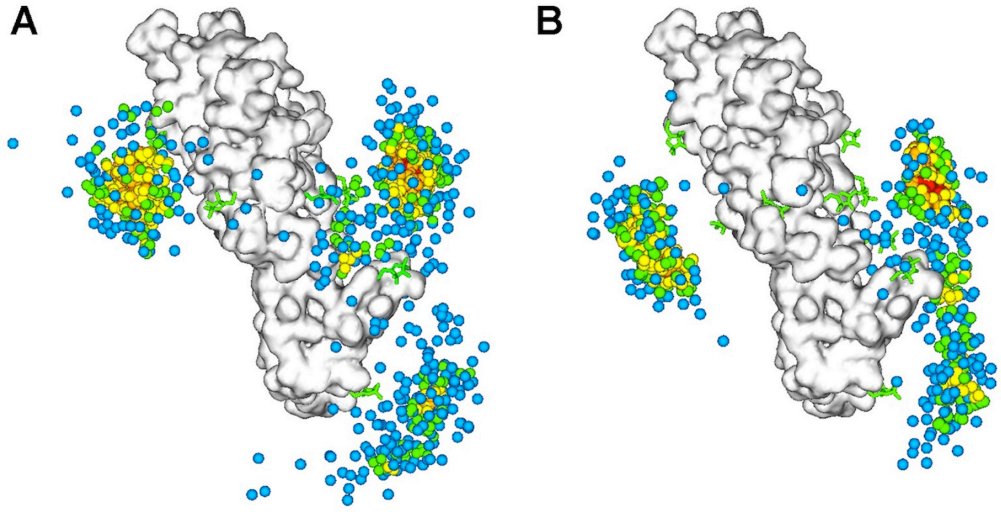

Figure 4.9. Comparison of the encounter complexes of $^{N\text{-}Ph}$complex (A) and $^{N\text{-}N}$complex (B). ^{N}Cyt *f* is shown as a white surface and spin labels as green sticks. Pc CoMs are represented by spheres, color-coded to indicate the density of the distributions, decreasing from red to blue. Densities were determined by counting the number of neighbours within 2.5 Å.

The centers-of-mass (CoMs) of Pc were colored according to the density of distribution, with red and blue representing the largest and smallest density, respectively. It should be noted that the incomplete coverage of spin labels on Cyt *f* surface implies that also other Cyt *f* surface areas could be involved in the encounter complex. The current analysis shows that the encounter complex is at least distributed over three extensive areas of Cyt *f* surface. All three encounters showed high density in the central regions and low density at the edge of the

encounter complex. The most extended area is located in the vicinity of the binding site found in the final complex models, the second is in front of the small domain of Cyt *f* and the third on the backside relative to the final complex. The third area is an artifact due to the use of the soluble part of ^{N}Cyt *f*. *In vivo*, Cyt *f* is embedded on the thylakoid membrane that will prevent Pc from binding on this side (Chapter III). In all three areas the interface comprises large patches of polar and hydrophobic residues. Despite the fact that in this study a less extensive portion of the Cyt *f* surface was monitored, the encounter complex resembles the one found for the $^{N\text{-}N}$complex (Figure 4.9B). The encounter ensemble of $^{N\text{-}Ph}$complex is more extensive and covers a larger area of the hydrophobic regions of Cyt *f*. In $^{N\text{-}N}$complex, stronger charge interactions may lead to more defined encounter regions. In the $^{N\text{-}N}$complex one continuous diffusive encounter region is present on the side of the binding site, while in the $^{N\text{-}Ph}$complex two distinct diffusive areas can be seen. To evaluate the distribution of the ET active complexes, the CoMs of ^{Ph}Pc are colored according the calculated distance between Cu in Pc and Fe in Cyt *f*, with red and blue representing the smallest and largest distance, respectively (Figure 4.10A). Like for the $^{N\text{-}N}$complex (Figure 4.10B), the encounter complex orientations compatible with rapid ET (red dots, Cu-Fe distance ≤ 16 Å) are located only in front of the haem, in the close vicinity of the binding site found in the final complex.

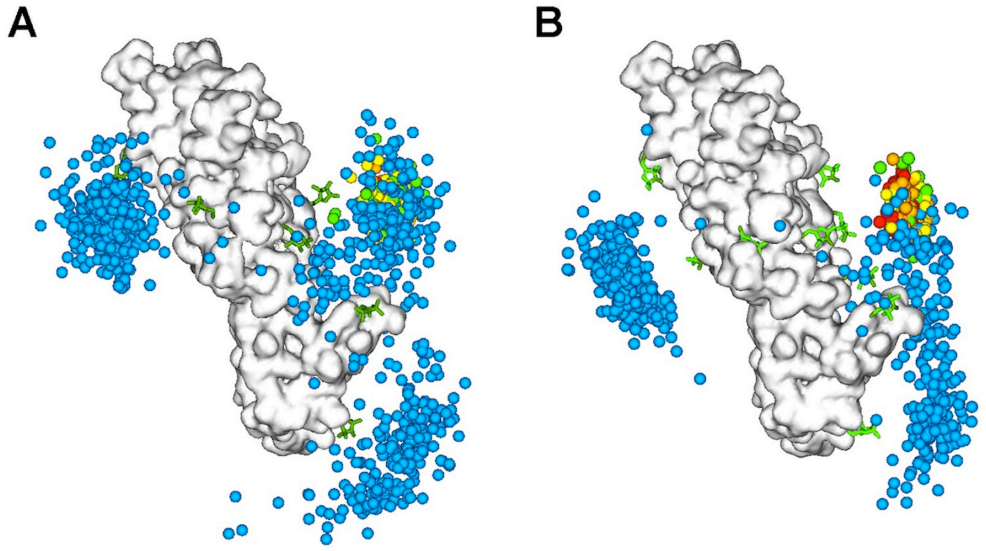

Figure 4.10. Comparison of the encounter complexes of $^{N\text{-}Ph}$complex (A) and $^{N\text{-}N}$complex (B). Cyt *f* is shown as a white surface and spin labels as green sticks. Pc CoMs are represented by spheres. Pc CoMs are color-coded to indicate the distance between Cu in Pc and Fe in Cyt *f*, increasing from red to blue (red ≤ 16 Å; orange ≤ 18 Å; yellow ≤ 20 Å; green ≤ 22 Å; blue > 22 Å).

Role of electrostatic interactions in complex formation

The effect of ionic strength (*I*) on the binding shifts of ^{Ph}Pc in the presence of reduced ^{N}Cyf *f* at Cyf *f*:Pc molar ratio 3:1 was investigated at NaCl concentrations of 100 mM (*I*= 110 mM) and 200 mM (*I*= 210 mM). The CSPs ($\Delta\delta_H$)

were defined relative to the control measurements recorded on free ^{Ph}Pc at the same ionic strenth values. The $\Delta\delta_H$ at the different salt concentrations were plotted versus ^{Ph}Pc residue numbers (Figure 4.11A).

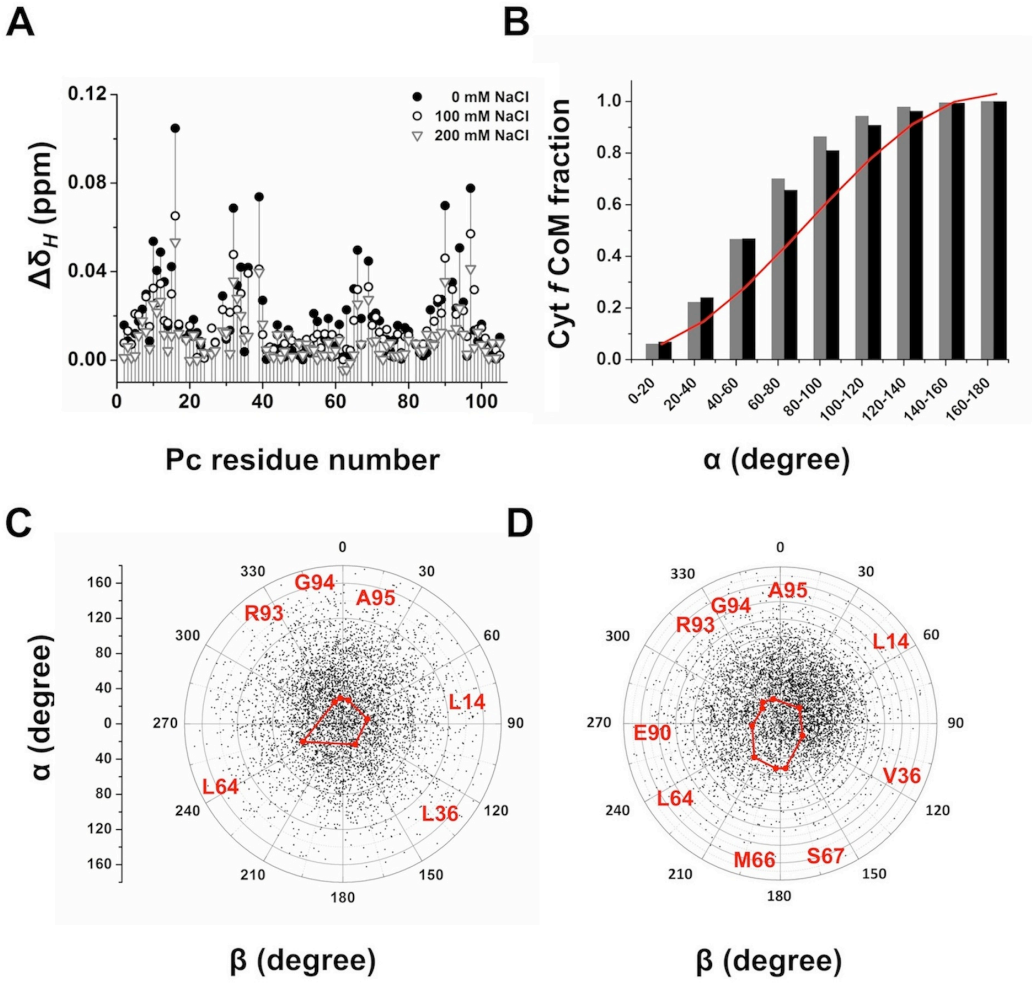

Figure 4.11. Role of electrostatic interactions in Cyt *f*-Pc complexes. A) Ionic strength dependence of $\Delta\delta_H$ for ^{Ph}Pc backbone amide protons in the presence of reduced ^{N}Cyt *f* at *I*= 10 mM (black dots), *I*= 110 mM (black circles) and *I*= 210 mM (grey triangles). (B, C, D) Analysis of the encounter complex generated by MC simulations. B) The cumulative fraction of Cyt *f* CoMs for the α angle is plotted for the $^{N\text{-}Ph}$complex (black bars) and $^{N\text{-}N}$complex (grey bars). The red line represents the cumulative fraction for a completely random distribution around a sphere. Plots of the position of the ^{N}Cyt *f* CoMs with respect to ^{Ph}Pc (C) and ^{N}Pc (D), in the MC ensembles. The red line connects the positions of hydrophobic patch residues. The Nε of H92 is at the center of the plots in panels C and D.

In rigid-body MC simulations the association of two proteins is simulated on the basis of their electrostatic potentials.[57] On the assumption that the formation of the encounter complex is purely driven by long-range electrostatic forces,[21] PRE and MC simulations were successfully combined for the visualization of the encounter complex of cytochrome *c* and cytochrome *c* peroxidase, demonstrating that the formation of this complex could be explained by

electrostatic interactions alone.[28] The same approach on the $^{N\text{-}N}$complex revealed to be inadequate to describe the encounter complex, which appears to be stabilized by electrostatic as well as hydrophobic interactions (Chapter III). At the same time, MC simulations provided evidence of the electrostatic preorientation of Pc towards Cyt *f*, as was found on the basis of CSP and PRE data. MC simulations were performed for the $^{N\text{-}Ph}$complex to establish whether electrostatic preorientation of ^{Ph}Pc can occur despite the negative charge of both proteins. The calculations produced an ensemble consisting of the Boltzmann distribution of orientations of Cyt *f* around Pc. An ensemble of 5000 structures was randomly selected from the entire set of two million solutions and the positions of Cyt *f* CoMs were plotted in Figure 4.11C. The position in the plot is determined by two angles. The first is the cone angle (α) formed by the Cyt *f* CoM, the Pc CoM and the Nε atom of copper ligand H92, taken as the center of the hydrophobic patch. The larger this angle is, the further the Cyt *f* CoM is rotated away from the hydrophobic patch. The α angle is represented by the circles in Figure 4.11C and 4.11D. The second angle, β, indicates the position on the cone, and represents the side of Pc to which the Cyt *f* CoM is rotated. The hydrophobic patch is delineated by a red line marked with residue numbers. Figure 4.11C shows that Cyt *f* binds in a diffusive manner, but toward the hydrophobic patch side of ^{Ph}Pc thantoward the other end. Clearly, preorientation occurs due to electrostatic interactions. This finding is also illustrated in Figure 4.11D. The cumulative fraction of Cyt *f* CoMs for the α angle is plotted (black bars). The red line represents the cumulative fraction for a completely random distribution around a sphere. The fraction of CoMs with α angles of less than 90° is larger than 50%, so more than half of the CoMs is present around the half of Pc that comprises the hydrophobic patch, due to electrostatic preorientation. This suggests that despite the net negative charge of ^{Ph}Pc, the localization of positive charges promotes the formation of and oriented complex. For comparison, the same calculations, on the basis of an earlier study (Chapter III) are shown for the $^{N\text{-}N}$complex in Figure 4.11B (grey bars) and 4.11D. For this complex, the preorientation is stronger and shows a more defined binding spot for α = 60°-80° and β = 30°-120°. The primary reason for this difference between the complexes of ^{N}Cyt *f* with ^{N}Pc and ^{Ph}Pc is the presence of two Lys residues (K11 and K20) in this region of ^{N}Pc, which are substituted by serine and asparagine, respectively, in ^{Ph}Pc.

To compare the importance of ionic strength on the formation of the different Cyt *f*-Pc complexes, MC simulations were performed for $^{N\text{-}N}$complex, $^{N\text{-}Ph}$complex and $^{Ph\text{-}Ph}$complex at ionic strength values of 10 mM, 110 mM and 210 mM (Figure 4.12A-C). In the cases of $^{N\text{-}Ph}$complex (Figure 4.12B) and $^{Ph\text{-}Ph}$complex (Figure 4.12C), very sparsely distributed encounter complexes were observed at higher values of *I*. For the $^{N\text{-}N}$complex (Figure 4.12A), though the increase in ionic strength resulted in the production of more diffusive encounters, in which the Cyt *f* distribution covers a wider area of Pc surface than observed at low ionic strength, a preferable docking area could be still recognized and related to the diverse electrostatic properties of ^{N}Pc. The histograms of the electrostatic interaction energies show that at an ionic strength of 210 mM (green bars), the $^{N\text{-}Ph}$complex (Figure 4.12B) and $^{Ph\text{-}Ph}$complex (Figure 4.12C) have lost all electrostatic attraction. For the $^{N\text{-}N}$complex (Figure 4.12A) it is strongly reduced but not completely zero.

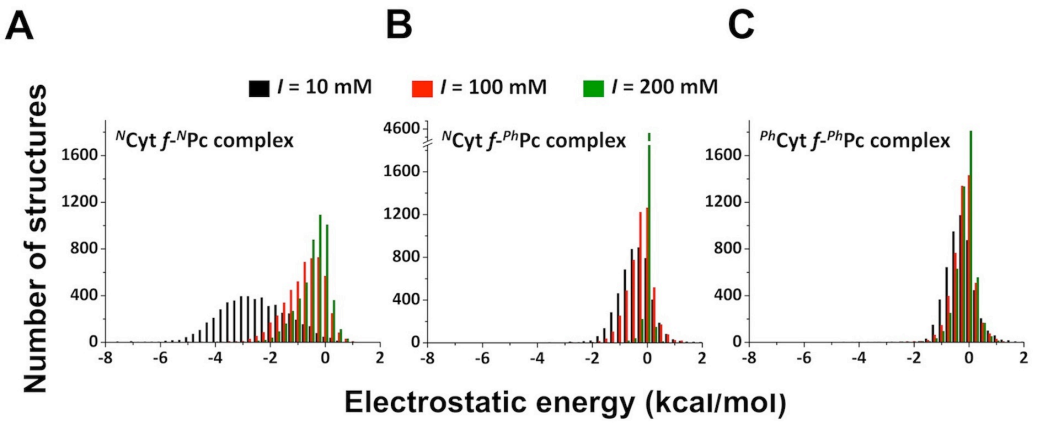

Figure 4.12. Electrostatic interaction histograms from MC simulation. MC simulations were performed at *I*= 10mM (black bars), *I*= 110mM (red bars) and *I*= 210mM (green bars) versus the electrostatic energy values for $^{N\text{-}N}$complexes (A), $^{N\text{-}Ph}$complexes (B) and $^{Ph\text{-}Ph}$complexes (C).

Comparison among Cyt *f*-Pc complexes

Recently, we proposed a model for the formation of the $^{N\text{-}N}$complex on the basis of the available kinetic and NMR data. Upon approach of the proteins, ^{N}Pc is rotated by electrostatic interactions to face ^{N}Cyt *f* with its hydrophobic patch leading to the formation of the encounter complex. This state is not stabilized only by charge interactions. Also hydrophobic interactions are important, allowing a smooth transition from encounter to ET-capable orientations by gradual increase of the hydrophobic overlap and sliding over the hydrophobic interface. It is interesting to interpret the data for the $^{N\text{-}Ph}$complex in the light of this model.
The most important difference between ^{N}Pc and ^{Ph}Pc is the net positive and negative charge, respectively. Given the highly negative charge on ^{N}Cyt *f* a poor interaction with ^{Ph}Pc is expected, if charge interactions are dominant. It was found that the affinity is five-fold lower for ^{Ph}Pc, suggesting that charges indeed play a role. This is also supported by the MC calculations that show less preorientation for ^{Ph}Pc than for ^{N}Pc. Nevertheless, some preorientation is still observed, indicating that the dipolar nature charge distribution is important in complex formation. The MC results are supported by the CSP and PCS data, which clearly demonstrate that the hydrophobic patch is the side of ^{Ph}Pc that is in contact with ^{N}Cyt *f*. However, the MC results do not agree quantitatively with the PRE data, indicating that electrostatic interactions alone are not sufficient to describe the encounter ensemble and the final complex.
The PCS based final complex shows predominantly hydrophobic contacts and the ^{Ph}Pc orientation is different from that in the $^{N\text{-}N}$complex, which can be explained by the substitution of several Lys residues on ^{Ph}Pc, resulting in the absence of several charge-charge interactions with negative residues on ^{N}Cyt f. The encounter complex produced using PRE driven ensemble docking is similar to that of the $^{N\text{-}N}$complex, though even more diffusive. In both encounter

complexes, Pc is found in contact with the non-polar surfaces of Cyt *f*, strongly suggesting that hydrophobic interactions indeed contribute to the encounter complex.

The PCS are much smaller in the $^{N\text{-}Ph}$complex than in the $^{N\text{-}N}$complex, suggesting that the encounter complex is more populated. The size of PCS strongly depends on the distance between the haem iron and the Pc nucleus that experiences the PCS. Thus, it is expected that in the encounter complex, which is spread over a large surface area of Cyt *f*, the PCS will be smaller than in the final complex. Orientation averaging may reduce the PCS further. The size of the PCS is about three-fold less for ^{Ph}Pc than for ^{N}Pc (Figure 4.3B).

In encounter complexes that are of an electrostatic nature CSPs are very small, compared to those in the final complex,[30,72] and increasing the fraction of the encounter complex strongly reduces the average size of the CSP in those complexes.[31] In complex with ^{N}Cyt *f* the CSPs for ^{Ph}Pc are also reduced compared to ^{N}Pc (Figure 4.1C, 4.1D) but not very much, much less than three-fold. This is an interesting observation because significant CSPs may be expected also in the encounter complex, if it is stabilized by hydrophobic contacts. The chemical shift of amide groups is particular sensitive to polarity and hydrogen bond formation, so the desolvation of the protein surface that accompanies the formation of hydrophobic contacts is expected to cause significant CSPs.

It is interesting to compare the effects of ionic strength in the $^{N\text{-}N}$complex, $^{N\text{-}Ph}$complex and $^{Ph\text{-}Ph}$complex. Addition of 200 mM NaCl to the $^{Ph\text{-}Ph}$complex (*I*= 210 mM) had essentially no effect on the fraction of bound Pc,[54] suggesting that hydrophobic contacts strongly dominate the interaction. The K_D was difficult to determine accurately and was reported to be about 1 mM. Here, we use a range of 1-3 mM. If it is assumed that the hydrophobic contribution to the binding is similar in the three complexes, the contribution of the electrostatic interactions can be estimated for the $^{N\text{-}N}$complex and the $^{N\text{-}Ph}$complex. An affinity of 1-3 mM equals a change in free energy of binding of 4.1 – 3.4 kcal/mol. The K_D values for the $^{N\text{-}N}$complex and the $^{N\text{-}Ph}$complex are 80[29] and 400 μM in the absence of salt (*I* = 10 mM), suggesting an additional contribution from the charge interactions of 1.5 – 2.2 kcal/mol and 0.55 – 1.2 kcal/mol, respectively. Thus, the electrostatic interaction represents 27%-38% and 12%-26% of the total binding energy in the $^{N\text{-}N}$complex and the $^{N\text{-}Ph}$complex.

Addition of 160 mM NaCl (*I* = 170 mM) to the $^{N\text{-}N}$complex reduced the fraction bound by about 50%[49] and it can be calculated on the basis of the protein concentrations used in that experiment that the binding energy decreased with 1.6 kcal/mol, nearly abolishing the charge-charge contribution. The same is observed for the $^{N\text{-}Ph}$complex, where addition of 200 mM NaCl (*I* = 210 mM) reduces the fraction bound by 60%, which translates to a loss of -0.8 kcal/mol of binding energy under the given experimental conditions. Thus, under the assumption that the hydrophobic contribution is conserved among these complexes, it can be concluded that the electrostatic contribution represents one-third of the binding energy for the $^{N\text{-}N}$complex at low ionic strength and much less at more physiological values. For the $^{N\text{-}Ph}$complex this fraction is even smaller. The trend is qualitatively supported by the electrostatic interaction histograms from the MC calculations (Figure 4.12A-C). This is an important finding in relation to earlier *in vivo* studies, in which no significant effects of mutation of charged

residues in the interface of an algal Cyt *f*-Pc complex could be detected in activity assay.[63-64] These result suggest that charge interactions are not relevant for the complex. On the other hand, the results on the cross-complex show that even weak electrostatic interactions are effective in pre-orientation Pc to face Cyt *f* with its hydrophobic patch. Furthermore, many charged residues on both proteins are conserved, especially among plants, suggesting that at least under some circumstances the charge interactions contribute significantly to the electron transfer process in photosynthesis.

In conclusion, the current study fully supports the model complex formation described for the $^{N\text{-}N}$complex. In $^{N\text{-}Ph}$complex the role of charges has not been abolished, but it is reduced in favor of hydrophobic contacts, creating a complex with biophysical properties that is a mixture of the $^{N\text{-}N}$complex and the $^{Ph\text{-}Ph}$complex. The variation that is observed between mechanisms of complex formation observed for the same complex from different species, shows that several ways exist to achieve both fast ET and rapid turn-over in protein complexes. The common denominator may be a low affinity and low energy barriers between the subsequent states in the reaction.

Chapter V

Concluding remarks

Many biological processes involve a balanced and highly organized network of specific proteins, which communicate with each other via physical interactions. Understanding the nature of these interactions is, therefore, a matter of primary importance. Cyt *f*-Pc complex is stabilized in its final orientation by both electrostatic and hydrophobic interactions. The aim of this thesis was to visualize the encounter state of Cyt *f*-Pc complex in order to understand the finely balanced hydrophobic and electrostatic interactions involved in the process of protein complex formation.

Encounter complex

The encounter complex model was initially introduced for small-molecules reactions[176,177] before being applied to macromolecular association theories.[178,179] In the diffusion-limited regime of protein association, the encounter complex can be defined as the intermediate formed after diffusion and subsequent interaction of the free proteins, which can potentially evolve to the final complex.[22] The formation of the encounter complex has long been considered to be mediated by long-range electrostatics. These interactions preserve the surface solvation of the individual proteins.[21] The model exhaustively describes complexes with electrostatic-assisted association and many examples have been reported.[23,28,30,171] The association between proteins with low charge complementarity could be theoretically described as driven by desolvation alone.[170] The addition of hydrophobic and electrostatic intercations to this model[71] indicated that the mechanism of association strictly depends on the surface properties of the specific proteins forming the complex. Our studies on the encounter state of *N-N*complex (Chapter III) and *N-Ph*complex (Chapter IV) provide novel experimental evidence that hydrophobic interactions participate to the formation and stabilization of the encounter complex as well as electrostatic interactions. In fact, the charge distribution on Pc results in the formation of a dipole that promotes the electrostatic pre-orientation of Pc with the hydrophobic patch towards the negatively charged Cyt *f*, allowing for the contact with the non-polar surfaces of Cyt *f* already in the encounter complex. The encounter and final state have similar energies and, therefore, the distinction between the encounter and the final complex seems to vanish. This new model for protein complex formation thus proposes that the encounter complex proceeds via a smooth and gradual transition to the final complex.

ET protein complexes

The inter-exchange between the two energetic states is highly consistent with the theory for which ET complexes exist in multiple active orientations.[20] In this regard, high dynamics were observed in several ET complexes.[28,30,72] The ET mainly depends on the distance between redox centers.[178] Once the redox centers reach a favorable distance from each other, ET can occur even between non-physiological redox partners[166] or between mutants of the physiological

partners.[7,8] The presence of multiple orientations seems to compensate for the low specificity by providing the necessary balance between association and dissociation required for efficient turnover in ET systems.[4] It is tempting to correlate the relative populations of encounter and final states to the biological function of the complex. In highly dynamic systems, such as ET complexes, the encounter state can be significantly populated[28] (Chapter III and IV) or the complex can even exist in a pure encounter state.[30] Instead, more specific systems, such as complexes involved in gene expression regulation,[23,141] seem to be also more static, having a greater tendency to prominently exist in a single, specific orientation, as required by the strict regulation to which these processes are subject.

Diamagnetic chemical shift perturbations analysis

NMR chemical shift perturbations (CSP) represent an extremely informative tool to study protein-protein interactions, especially for weak and dynamic complexes, such as that of Pc and Cyt *f*. In a typical experiment, the HSQC spectrum of ^{15}N-labeled Pc is monitored and the perturbations of the chemical shifts were recorded while increasing concentration of unlabeled Cyt *f*. The interaction with Cyt *f* causes changes on the surface of Pc, which affect the chemical shifts of the amide nuclei in the area involved in the complex formation, providing a residue-resolution map of the binding interface (Chapter IV). Titrations of Cyt *f* into Pc also provide a convenient way to establish the affinity and specificity of binding. From the CSP analysis it also was possible to establish the interference of spin labels on the *N-N*complex formation (Chapter II and III).

The degree of variability observed in the size of the chemical shifts among different complexes demonstrates to be a qualitative indication for intra-complex dynamics in electrostatic complexes.[15,31,32,72,73] A complex existing predominantly in a single orientation will yield large binding shifts as a result of desolvation and formation of specific, short-range interactions. In highly dynamic complexes, a significant fraction of the total population is present in the encounter state, consisting of multiple orientations that reduce the binding shifts by averaging.[9] Thus, the decrease of the average size of CSP in the *N-Ph*complex compared to those of *N-N*complex, supported the higher degree of dynamics suggested by PRE experiments and docking simulations (Chapter IV). This work has provided an interesting new observation, namely that the absolute CSP size observed in purely electrostatic complexes[30,31,72,73] is much smaller than that observed for *N-N*complex and *N-Ph*complex, which have shown to be dominated by hydrophobic as well as electrostatic interactions. The presence of hydrophobic interactions in the encounter complex implies the removal of water molecules that results in a big binding shift. The decrease of the CSP caused by dynamics is thus partially compensated by the effect of the desolvation of the encounter binding surfaces.

Paramagnetic NMR

The presence of a paramagnetic center in a protein can be a valuable source of structural information by affecting the chemical shifts and the nuclear relaxation rates of the observed nucleus in a distance-dependent manner.[180] The possibility to introduce a paramagnetic centre via site-directed labeling has extended the application of paramagnetic NMR vastly. The pseudocontact shifts caused by the intrinsic paramagnetic haem iron of ^{N}Cyt *f*, provides sufficient restraints for the determination of an ensemble of orientations of ^{Ph}Pc in the final $^{N\text{-}Ph}$complex (Chapter IV). The size of PCS is strongly related to the distance between the observed nucleus and the iron and depends also on the magnetic susceptibility anisotropy tensor. As a result, the presence of multiple orientations in a complex will reduce the average PCS, yielding a qualitative correlation between the PCS size and the degree of dynamics within the complex.[32] Indeed, the in $^{N\text{-}Ph}$complex the haem induces smaller PCS than in the $^{N\text{-}N}$complex, similarly to what was observed for the CSPs. Both reflect the larger dynamics within the cross complex (Chapter IV).
Paramagnetic relaxation enhancement is a well-established method for classical structural determination studies.[181] The sensitivity of PRE to lowly populated states makes it a versatile technique to investigate dynamic processes involved in complex formation.[96] The attachment of spin labels at several locations on the Cyt *f* surface allowed for the detection of the diffusive encounter states of $^{N\text{-}N}$complex and $^{N\text{-}Ph}$complex to be characterized (Chapter III and IV, respectively). The PRE patterns clearly indicated that both ^{N}Pc and ^{Ph}Pc are pre-oriented towards ^{N}Cyt *f* during the search of the specific binding site. The accurate maps of the binding interface of the encounter complexes obtained from PREs provided residue-resolution information on the region directly explored by the encounter complex.
To characterize the entire complex, both encounter and final states, a single method is inadequate. PRE and PCS provide complementary views that together help to obtain a more accurate picture of the complex.

Computational methods to study protein interactions

The field of structural biology is receiving immense benefits from the parallel development of computational methodologies to predict protein docking. In a nutshell, these methods use the coordinates of the unbound proteins to obtain computationally a model of the bound complex on the basis of either experimental data[182] or theoretical assumptions.[183-185]
Structural information, as gained by X-ray crystallography or NMR spectroscopy, can be explicitly considered and treated as active and driving force for the docking.[154,186] Recent advances in the use of paramagnetic NMR data[174] allow new data-driven docking techniques to determine the solution structures, even of large molecules and complexes[187,188] and transient protein-protein complexes.[26] NMR restraint-guided docking has been used to determine the final orientation of $^{N\text{-}N}$complex (Chapter II) and $^{N\text{-}Ph}$complex (Chapter IV) on the basis of observed

PREs and PCSs, respectively. Because many protein complexes exist and function as a dynamic ensemble, there is a growing need to be able to model such ensembles even though the interpretation of experimental data is not straightforward. The challenge of visualizing a dynamic encounter complex on the basis of experimental PREs was elegantly addressed by Prof. Clore and coworkers by representing one of the interaction partner as an ensemble of conformers, which is docked simultaneously to the single copy of the other protein.[141] This approach was thus used for the visualization of the encounter state on the basis of the experimental PREs both for the $^{N\text{-}N}$complex and $^{N\text{-}Ph}$complex (Chapter III and IV, respectively). In both encounter complexes, Pc was found in contact with the non-polar surfaces of Cyt *f*, strongly suggesting that hydrophobic interactions indeed contribute to the encounter complex. The encounter complexes thus produced using PRE driven ensemble docking were similar for location and distribution. The encounter of the $^{N\text{-}Ph}$complex resulted to have a more diffuse nature, reflecting the higher degree of dynamics with respect to the $^{N\text{-}N}$complex. Because the observed PRE is a population weighted average of all species present in solution,[84] many possible docking solutions can correspond to the observed PREs, limitating thus the accuracy of the method to furnish a high-resolution picture of the encounter complex.

Softwares for theoretical prediction of protein complex structure and association are mainly based on shape complementarity,[189] electrostatics[57,109] and solvation terms.[190] The primary methods for the computational study of protein association, such as BD and MC, simulate the complex formation on the basis of the electrostatic properties of the individual proteins. On the assumption that the electrostatic forces dominate the encounter complex formation, MC simulations were used to obtain a structural description of the encounter complex of $^{N\text{-}N}$complex (Chapter III) and $^{N\text{-}Ph}$complex (Chapter IV). MC approach did not produce encounter complexes in agreement with the experimental PREs, indicating that electrostatic interactions are not dominant in these Cyt *f*-Pc complexes. Still, MC simulations also provided evidence for the electrostatic pre-orientation of Pc in both complexes, in qualitative agreement with CSP and PRE interaction maps. Thus, the MC approach was confirmed to be a powerful tool to evaluate the contribution of electrostatic forces in complex formation.

Nederlandse Samenvatting

De levende cel is een open en dynamisch systeem dat zich voortdurend aanpast als gevolg van interne en externe prikkels. De *steady state* van de cel wordt in stand gehouden dankzij een efficiënt evenwicht van verscheidene metabole routes, waarbij veelal interacties tussen eiwitten een rol spelen. De eerste stap van eiwitassociatie is de vorming van een tijdelijk intermediair, gedefinieerd als het *encounter complex*, dat vervolgens ofwel dissocieert, ofwel overgaat in het productieve eindcomplex. De vorming van het *encounter complex* is een gevolg van Brownse diffusie en elektrostatische interacties die werken over lange afstanden, terwijl in het eindcomplex interacties over korte afstand plaatsvinden. Het eindcomplex is dus specifiek, door selectieve herkenning van de partner als gevolg van de oppervlakte-eigenschappen van individuele eiwitten. Om de mechanismen te begrijpen die de moleculaire herkenning reguleren, is het complex bestudeerd dat gevormd wordt door Cyt *f* en Pc van de cyanobacterie *Nostoc*. Deze eiwitten zijn redoxpartners in de zuurstofgenererende fotosynthese van planten, groene algen en cyanobacteriën. Het complex speelt een rol in de elektronoverdracht en is kortlevend door de lage specificiteit en de lage complementariteit van de oppervlaktes van Cyt *f* en Pc. Het kortlevende karakter van het complex is gerelateerd aan de hoge K_D die voor verschillende complexen van Cyt *f*-Pc zijn gemeten met kinetische methoden en kernspinresonantie. Karakteristiek voor zowel Cyt *f* als Pc uit alle organismen is de aanwezigheid van een uitgebreid hydrofoob oppervlak rond de redoxcentra. Daarentegen variëren de elektrostatische oppervlakte-eigenschappen aanzienlijk tussen de eiwitten uit verschillende organismen en lijken ze de oriëntatie van Pc in het eindcomplex te bepalen, evenals de mate van dynamiek binnen de complexen. In het geval van *Nostoc* hebben Cyt *f* en Pc als geheel respectievelijk een negatieve en positieve lading. Kernspinresonantie en theoretische simulaties zijn de belangrijkste technieken van het onderzoek dat in dit proefschrift beschreven staat. Deze technieken maken het mogelijk om de structurele aspecten van de Cyt *f*-Pc complexvorming te beschrijven.

Voor het eerst is de interactie tussen Cyt *f* en Pc van *Nostoc* onderzocht met behulp van *paramagnetic relaxation enhancement* (PRE). Drie posities rondom de veronderstelde bindingsplaats zijn geselecteerd voor mutatie naar cysteïne voor de bevestiging van spin labels. De posities van de mutaties zijn gekozen op basis van de structuur van het wild type complex in oplossing. Hierbij was gebruik gemaakt van *pseudocontact shifts* (PCS) van de resonanties van Pc kernen, veroorzaakt door de heem van Cyt *f*. De PRE-analyse laat zien dat het complex zeer dynamisch is, wat suggereert dat het *encounter complex* een belangrijke fractie van het complex vormt. Deze eerste resultaten suggereerden dat het complex in meerdere oriëntaties voorkomt. Bovendien toonden zij de noodzaak aan om meer inzicht te krijgen in de verdeling van het *encounter complex* over het oppervlak van Cyt *f* om een volledige beschrijving van de associatie van Cyt *f* en Pc te kunnen verkrijgen.

Voor een betere beschrijving van het *encounter complex* is de aanvankelijke set uitgebreid tot negen spin labels, goed verdeeld over het oppervlak van Cyt *f*. De metingen en analyses van PRE's laten zien dat Pc een groot deel van het oppervlak van Cyt *f* aftast. De overeenkomst tussen de waargenomen PRE-patronen voor Pc kernen voor spin labels die dichtbij de heem van Cyt *f* zijn bevestigd geeft aan dat Pc met één zijde het oppervlak van Cyt *f* aftast.

Elektrostatische interacties pre-oriënteren vermoedelijk het hydrofobe oppervlak van Pc naar Cyt *f* en faciliteren het contact van de eiwitten, zodat het *encounter complex* wordt gevormd. Vervolgens wordt dit complex door hydrofobe interacties gestabiliseerd. De waargenomen PRE-data in het regime van snelle uitwisseling zijn populatie-gewogen gemiddeldes van de PRE's voor het complex Cyt *f*-Pc in zowel het *encounter complex* als eindcomplex. De PRE verschaft daarom alleen kwalitatieve informatie over het *encounter complex*. De visuele weergave van het *encounter complex* is verkregen met behulp van theoretische modellen. Het programma MC-DOCK simuleert de complexvorming enkel op grond van de elektrostatische interacties tussen de eiwitten. Het *encounter complex* wordt dan ook voorgesteld als een Boltzmanndistributie van complexen op grond van de elektrostatische interactie-energie. Deze computersimulaties leverden een model van het Cyt *f*-Pc *encounter complex* op dat niet volledig in overeenstemming was met de experimentele PRE's. Dit geeft aan dat de vorming van het *encounter complex* niet kan worden beschreven door enkel elektrostatische interacties. Een tweede methode die is toegepast is de *ensemble docking* waarbij enkele kopieën van Pc tegelijk kunnen binden aan Cyt *f* op grond van de PRE-data. Het resulterende *encounter complex* vertegenwoordigt een waarschijnlijkheidsverdeling van oriëntaties van Pc in het complex op grond van de experimentele gegevens. Dit *encounter complex* heeft een diffuus karakter, waarin Pc en Cyt *f* contact maken met hun hydrofobe oppervlaktes. Ladingsinteracties dragen dus voornamelijk bij aan de pre-oriëntatie van Pc naar Cyt *f*, terwijl de hydrofobe interacties zorgen voor de vorming en stabilisatie van het *encounter complex*. Het diffuse karakter van het *encounter complex* suggereert dat voor dit systeem een goed gedefinieerde eindoriëntatie niet een fundamentele voorwaarde is voor de elektronoverdracht. Een efficiënt en kortlevend complex is noodzakelijk voor snelle elektronoverdracht door de gehele fotosynthetische redoxketen. Dat sluit zelfs de vorming van een hecht complex uit want daarvan duurt de dissociatie te lang. De fotosynthese is dus gebaat bij een complex dat functioneel is in meerdere oriëntaties. De dubbelrol van hydrofobe interacties voor complexvorming en de stabilisatie van het eindcomplex suggereert een lage energiebarrière van *encounter complex* naar eindcomplex en een soepele overgang tussen de twee energietoestanden. Op basis van deze waarnemingen is een nieuw model voor eiwitassociatie voorgesteld waarin de vorming van het *encounter complex* gedurende alle fases van de associatie een vlak energielandschap is, zonder een helder onderscheid tussen het *encounter complex* en eindcomplex.

Er werd een tussenliggende affiniteit geconstateerd bij het niet-fysiologische complex van Cyt *f* van *Nostoc* met Pc van *Phormidium laminosum*, vergeleken bij de fysiologische complexen van *Nostoc* en *P. laminosum*. ^{Ph}Pc kan worden beschouwd als een natuurlijke variant op ^{N}Pc waarvan de gehele lading negatief is. Het complex laat een "head-on"-oriëntatie zien die lijkt op die van het complex van *Phormidium*, waarin enkel het hydrofobe stuk van Pc contact maakt met het hydrofobe stuk van Cyt *f* rondom de heem. De verminderde elektrostatische aantrekking leidt tot een meer diffuse verdeling van het *encounter complex* dan in *Nostoc* en suggereert een hogere mate van dynamiek voor het "kruiscomplex". Zouttitraties en Monte-Carlo computersimulaties laten zien dat elektrostatische pre-oriëntatie toch voorkomt in

dit complex en bijdraagt aan de associatie, ondanks het feit dat ^{N}Cyt *f* en ^{Ph}Pc beide negatief zijn. Deze bevindingen ondersteunen het nieuw voorgestelde model voor complexvorming, waarin hydrofobe en elektrostatische interacties samen de associatie van Cyt *f* en Pc bevorderen.

English Summary

The living cell is an open and dynamic system that continuously adapts to both inner and external stimuli. The cellular steady state is maintained thanks to the efficient balance between the different biochemical pathways, most of which involve interactions between proteins. Protein association requires the initial formation of a transient intermediate, defined as encounter complex, which can either dissociate or evolve to a final, productive complex. The formation of the encounter complex was described as resulting from Brownian diffusion and long-range electrostatics, whereas short-range interactions take place in the final complex. Thus, a productive protein complex implies the selective recognition between the specific interaction partners, determined by the surface properties of the individual proteins. In order to understand the mechanisms that control molecular recognition the complex formed by Cyt *f* and Pc from the cyanobacterium *Nostoc* was studied. These proteins are redox partners within the oxygenic photosynthetic chain in plants, green algae and cyanobacteria. As a consequence of the electron transfer function, the complex shows a transient nature, resulting from low specificity and surface complementarity between Cyt *f* and Pc. Another important transient feature of the complex is represented by the high K_D measured in several Cyt *f*-Pc complexes, which have been characterized by either kinetic or NMR studies. The presence of an extended hydrophobic patch surrounding the redox centers is a common feature to both Cyt *f* and Pc among different organisms. On the contrary, the electrostatic surface properties significantly vary between different organisms and seem to influence the final orientation and the degree of dynamics within the complexes. In the particular case of *Nostoc*, Cyt *f* and Pc have an overall negative and positive charge, respectively. The key techniques used in this thesis are NMR spectroscopy and computational methods, which allowed for the description of both the dynamic and structural aspects of Cyt *f*-Pc complex formation.

For the first time, the interaction between Cyt *f* and Pc from *Nostoc* was investigated by paramagnetic relaxation enhancement (PRE). Three sites around the putative binding site for ET on Cyt *f* were selected for spin label attachment. The positions of the mutations were designed on the basis of the solution structure of the wild type complex, as determined by taking advantage of the pseudocontact shifts (PCS) generated by the haem of Cyt *f* on Pc nuclei. It was found that the complex is highly dynamic, suggesting a significant population of the encounter complex. These early results suggested the existence of the complex in multiple orientations and, consequently, also indicated the need to understand the distribution of the encounter complex to provide a complete description of the association between Cyt *f* and Pc.

For the detection of Cyt *f*-Pc encounter complex, the initial data set was extended to nine spin labels, distributed over a wide area on Cyt *f* surface. The measurements of PREs and their analysis showed that Pc samples an extended portion of Cyt *f* surface. The similarity between the PRE patterns observed in the presence of spin labels attached near to the haem of Cyt *f* indicated that Pc samples the Cyt *f* surface with a single patch. Electrostatic interactions are thought to pre-orient Pc with the hydrophobic patch towards Cyt *f* and to favor the contact of the proteins to form the encounter complex, which is then stabilized by hydrophobic interactions. The observed PRE data in the fast exchange regime are population weighted averages of the PREs for Cyt *f*-Pc

complex in both the encounter and final states. The observed PRE therefore provide only qualitative information about the encounter complex. The visual representation of the encounter complex was obtained by using theoretical models. MC-dock predicts the formation of a complex solely on the basis of the electrostatic surface properties of the interacting proteins. The encounter complex is represented by a Boltzmann distribution of complexes according to their electrostatic-interaction energy. MC simulations of Cyt *f*-Pc encounter complex did not provide a model in agreement with the experimental PREs, indicating that electrostatic interactions alone cannot describe the formation of this encounter complex. In the ensemble docking method, the driving force for complex formation is given by the observed PREs. The resulting encounter complex represents a probability distribution of Cyt *f*-Pc orientations, which account for the experimental data. This encounter complex has a diffusive nature, in which Pc can diffuse on the non-polar region of Cyt *f* by overlap with its own hydrophobic patch. Long-range electrostatics mainly contribute to the pre-orientation of Pc towards Cyt *f*, as indicated by the PRE analysis, and hydrophobic interactions in the formation and stabilization of the encounter complex. The diffuse nature of the encounter complex suggests that in this system a final, well-defined orientation of the complex could be not a fundamental requirement for the ET function. In fact, the efficient turnover required for rapid ET through the photosynthetic redox chain precludes the formation of a tight complex and favors the existence of an ET active complex in multiple orientations. The dual role of hydrophobic interactions either in the formation of the encounter complex and in the stabilization of the final complex suggests a small energy barrier between the encounter and the final complex, favoring the smooth transition between the two energetic states. In the light of these observations, a new model to describe protein association has been proposed, in which the formation of the encounter complex presents a relatively flat energy landscape during all phases of the association, without a clear distinction between the encounter and the final complex.

The non-physiological complex formed by *Nostoc* Cyt *f* and *Phormidium laminosum* Pc was found to have an intermediate affinity between the physiological complexes of *Nostoc* and *Phormidium laminosum*. ^{Ph}Pc can be considered as a natural variant of ^{N}Pc, in which the overall charge is negative. The complex showed a "head-on" orientation, reminiscent of that found in the *Phormidium* complex, in which only the hydrophobic patch of Pc makes contact with the hydrophobic region surrounding the haem on Cyt *f*. The reduced electrostatic attraction also seems to favor a more diffusive distribution of the encounter complex than in *Nostoc*, suggesting a higher degree of dynamics within the cross-complex. Interestingly, despite the reduced electrostatic attraction between ^{N}Cyt *f* and ^{Ph}Pc, salt-titration experiments and MC simulations showed that electrostatic pre-orientation is still occurring and contributes to the association. These findings supported the new model proposed for protein complex formation, in which hydrophobic and electrostatic interactions together promote the association of Cyt *f* and Pc.

Reference List

1. Northrup, S. H.; Erickson, H. P. *Proc. Natl. Acad. Sci. U. S. A* **1992,** *89*, 3338-3342.
2. Shapiro, R.; Vallee, B. L. *Biochemistry* **1991,** *30*, 2246-2255.
3. Albeck, S.; Schreiber, G. *Biochemistry* **1999,** *38*, 11-21.
4. Crowley, P. B.; Ubbink, M. *Acc. Chem. Res.* **2003,** *36*, 723-730.
5. Kang, C. H.; Ferguson-Miller, S.; Margoliash, E. *J. Biol. Chem.* **1977,** *252*, 919-926.
6. Setif, P. Q.; Bottin, H. *Biochemistry* **1995,** *34*, 9059-9070.
7. Schlarb-Ridley, B. G.; Bendall, D. S.; Howe, C. J. *Biochemistry* **2002,** *41*, 3279-3285.
8. Albarran, C.; Navarro, J. A.; Molina-Heredia, F. P.; Murdoch, P. S.; De la Rosa, M. A.; Hervas, M. *Biochemistry* **2005,** *44*, 11601-11607.
9. Prudencio, M.; Ubbink, M. *J. Mol. Recognit.* **2004,** *17*, 524-539.
10. Schreiber, G.; Fersht, A. R. *J. Mol. Biol.* **1995,** *248*, 478-486.
11. Clackson, T.; Wells, J. A. *Science* **1995,** *267*, 383-386.
12. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2007,** *68*, 803-812.
13. Crowley, P. B.; Carrondo, M. A. *Proteins* **2004,** *55*, 603-612.
14. Kiel, C.; Selzer, T.; Shaul, Y.; Schreiber, G.; Herrmann, C. *Proc Natl Acad Sci USA* **2004,** *101*, 9223-9228.
15. Volkov, A. N.; Bashir, Q.; Worrall, J. A. R.; Ubbink, M. *J. Mol. Biol.* **2009,** *385*, 1003-1013.
16. Schreiber, G.; Keating, A. E. *Curr. Opin. Struct. Biol.* **2011,** *21*, 50-61.
17. Williams, P. A.; Fulop, V.; Leung, Y. C.; Chan, C.; Moir, J. W.; Howlett, G.; Ferguson, S. J.; Radford, S. E.; Hajdu, J. *Nat. Struct. Biol.* **1995,** *2*, 975-982.
18. De Rienzo. F.; Gabdoulline, R. R.; Menziani, M. C.; Wade, R. C. *Protein Sci.* **2000,** *9*, 1439-1454.
19. Nersissian, A. M.; Shipp, E. L. *Adv. Protein Chem.* **2002,** *60*, 271-340.
20. Bashir, Q.; Scanu, S.; Ubbink, M. *FEBS J.* **2011,** *278*, 1391-1400.
21. Ubbink, M. *FEBS Lett.* **2009,** *583*, 1060-1066.
22. Schreiber, G. *Curr. Opin. Struct. Biol.* **2002,** *12*, 41-47.
23. Schreiber, G.; Fersht, A. R. *Nat. Struct. Biol.* **1996,** *3*, 427-431.
24. Harel, M.; Spaar, A.; Schreiber, G. *Biophys. J.* **2009,** *96*, 4237-4248.

25. Keizers, P. H. J.; Ubbink, M. *Prog. Nucl. Mag. Res. Sp.* **2011,** *58*, 88-96.

26. Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc Natl Acad Sci USA* **2006,** *103*, 18945-18950.

27. Kim, Y. C.; Tang, C.; Clore, G. M.; Hummer, G. *Proc. Natl. Acad. Sci. U. S. A* **2008,** *105*, 12855-12860.

28. Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010,** *132*, 241-247.

29. Scanu, S.; Forster, J.; Finiguerra, M. G.; Shabestari, M. H.; Huber, M.; Ubbink, M. *Chembiochem* **2012**.

30. Xu, X. F.; Reinle, W. G.; Hannemann, F.; Konarev, P. V.; Svergun, D. I.; Bernhardt, R.; Ubbink, M. *J. Am. Chem. Soc.* **2008,** *130*, 6395-6403.

31. Volkov, A. N.; Bashir, Q.; Worrall, J. A. R.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010,** *132*, 11487-11495.

32. Hulsker, R.; Baranova, M. V.; Bullerjahn, G. S.; Ubbink, M. *J. Am. Chem. Soc.* **2008,** *130*, 1985-1991.

33. Mitchell, P. *FEBS Lett.* **1977,** *78*, 1-20.

34. Hervas, M.; Navarro, J. A.; Diaz, A.; Bottin, H.; De la Rosa, M. A. *Biochemistry* **1995,** *34*, 11321-11326.

35. Mason, J. M.; Bendall, D. S.; Howe, C. J.; Worrall, J. A. *Biochim. Biophys. Acta* **2012,** *1824*, 311-318.

36. Worrall, J. A.; Luisi, B. F.; Schlarb-Ridley, B. G.; Bendall, D. S.; Howe, C. J. *Biochem. Soc. Trans.* **2008,** *36*, 1175-1179.

37. Howe, C. J.; Schlarb-Ridley, B. G.; Wastl, J.; Purton, S.; Bendall, D. S. *J. Exp. Bot.* **2006,** *57*, 13-22.

38. Katoh, S. *Nature* **1960,** *186*, 533-534.

39. Gorman, D. S.; Levine, R. P. *Proc. Natl. Acad. Sci. U. S. A* **1965,** *54*, 1665-1669.

40. Canters, G. W.; Gilardi, G. *FEBS Lett.* **1993,** *325*, 39-48.

41. Gray, H. B.; Malmstrom, B. G.; Williams, R. J. *J. Biol. Inorg. Chem.* **2000,** *5*, 551-559.

42. Schmidt, L.; Christensen, H. E.; Harris, P. *Acta Crystallogr. , Sect. D:Biol. Crystallogr.* **2006,** *62*, 1022-1029.

43. Bertini, I.; Ciurli, S.; Dikiy, A.; Fernàndez C.O.; Luchinat, C.; Safarov, N.; Shumilin, S.; Vila, A.J. *J. Am. Chem. Soc.* **2001,** *123*, 2405-2413.

44. Adman, E. T. *Adv. Protein Chem.* **1991,** *42*, 145-197.

45. Bertini, I.; Bryant, D.A.; Ciurli, S.; Dikiy, A.; Fernàndez C.O.; Luchinat, C.; Safarov, N.; Vila, A.J.; Zhao J. *J. Biol. Chem.* **2001,** *276*, 47217-47226.

46. Ubbink, M.; Ejdeback, M.; Karlsson, B. G.; Bendall, D. S. *Structure* **1998,** *6*, 323-335.

47. Ubbink, M.; Ejdeback, M.; Crowley, P. B.; Schlarb, B. G.; Howe, C. J.; Karlsson, B. G.; Bendall, D. S.; Canters, G. W. *J. Inorg. Biochem.* **1999,** *74*, 321.

48. Ejdeback, M.; Bergkvist, A.; Karlsson, B. G.; Ubbink, M. *Biochemistry* **2000,** *39*, 5022-5027.

49. Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A.; Crowley, P. B.; Ubbink, M. *Biochemistry* **2005,** *44*, 3176-3183.

50. Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A.; Ubbink, M. *J. Biol. Chem.* **2005,** *280*, 18908-18915.

51. Cramer, W. A.; Soriano, G. M.; Ponomarev, M.; Huang, D.; Zhang, H.; Martinez, S. E.; Smith, J. L. *Annu. Rev. Plant Physiol Plant Mol. Biol.* **1996,** *47*, 477-508.

52. Stroebel, D.; Choquet, Y.; Popot, J. L.; Picot, D. *Nature* **2003,** *426*, 413-418.

53. Martinez, S. E.; Huang, D.; Szczepaniak, A.; Cramer, W. A.; Smith, J. L. *Structure* **1994,** *2*, 95-105.

54. Crowley, P. B.; Otting, G.; Schlarb-Ridley, B. G.; Canters, G. W.; Ubbink, M. *J. Am. Chem. Soc.* **2001,** *123*, 10444-10453.

55. Cruz-Gallardo, I.; Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A. *FEBS Lett.* **2012,** *586*, 646-652.

56. Pearson, D. C., Jr.; Gross, E. L.; David, E. S. *Biophys. J.* **1996,** *71*, 64-76.

57. Ullmann, G. M.; Knapp, E. W.; Kostic, N. M. *J. Am. Chem. Soc.* **1997,** *119*, 42-52.

58. Pearson, D. C., Jr.; Gross, E. L. *Biophys. J.* **1998,** *75*, 2698-2711.

59. Kannt, A.; Young, S.; Bendall, D. S. *Biochim. Biophys. Acta* **1996,** *1277*, 115-126.

60. Gong, X. S.; Wen, J. Q.; Fisher, N. E.; Young, S.; Howe, C. J.; Bendall, D. S.; Gray, J. C. *Eur. J. Biochem.* **2000,** *267*, 3461-3468.

61. De Rienzo. F.; Gabdoulline, R. R.; Menziani, M. C.; De Benedetti, P. G.; Wade, R. C. *Biophys. J.* **2001,** *81*, 3090-3104.

62. Lange, C.; Cornvik, T.; Diaz-Moreno, I.; Ubbink, M. *Biochim. Biophys. Acta* **2005,** *1707*, 179-188.

63. Soriano, G. M.; Ponamarev, M. V.; Tae, G. S.; Cramer, W. A. *Biochemistry* **1996,** *35*, 14590-14598.

64. Soriano, G. M.; Ponamarev, M. V.; Piskorowski, R. A.; Cramer, W. A. *Biochemistry* **1998,** *37*, 15120-15128.

65. Meyer, T. E.; Zhao, Z. G.; Cusanovich, M. A.; Tollin, G. *Biochemistry* **1993,** *32*, 4552-4559.

66. Albarran, C.; Navarro, J.; De la Rosa, M. A.; Hervas, M. *Biochemistry* **2007,** *46*, 997-1003.

67. Schlarb-Ridley, B. G.; Navarro, J. A.; Spencer, M.; Bendall, D. S.; Hervas, M.; Howe, C. J.; De la Rosa, M. A. *Eur. J. Biochem.* **2002,** *269*, 5893-5902.

68. Gross, E. L. *Biophys. J.* **2004,** *87*, 2043-2059.

69. Hart, S. E.; Schlarb-Ridley, B. G.; Delon, C.; Bendall, D. S.; Howe, C. J. *Biochemistry* **2003,** *42*, 4829-4836.

70. Schlarb-Ridley, B. G.; Bendall, D. S.; Howe, C. J. *Biochemistry* **2003,** *42*, 4057-4063.

71. Gabdoulline, R. R.; Wade, R. C. *J. Am. Chem. Soc.* **2009,** *131*, 9230-9238.

72. Worrall, J. A. R.; Liu, Y. J.; Crowley, P. B.; Nocek, J. M.; Hoffman, B. M.; Ubbink, M. *Biochemistry* **2002,** *41*, 11721-11730.

73. Worrall, J. A. R.; Reinle, W. G.; Bernhardt, R.; Ubbink, M. *Biochemistry* **2003,** *42*, 7068-7076.

74. Worrall, J. A. R.; Kolczak, U.; Canters, G. W.; Ubbink, M. *Biochemistry* **2001,** *40*, 7069-7076.

75. Iwahara, J.; Anderson, D. E.; Murphy, E. C.; Clore, G. M. *J. Am. Chem. Soc.* **2003,** *125*, 6634-6635.

76. Prudencio, M.; Rohovec, J.; Peters, J. A.; Tocheva, E.; Boulanger, M. J.; Murphy, M. E. P.; Hupkes, H. J.; Kosters, W.; Impagliazzo, A.; Ubbink, M. *Chemistry-A European Journal* **2004,** *10*, 3252-3260.

77. Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004,** *126*, 12800-12808.

78. Vlasie, M. D.; Comuzzi, C.; van den Nieuwendijk, A. M. C. H.; Prudencio, M.; Overhand, M.; Ubbink, M. *Chem- Eur. J.* **2007,** *13*, 1715-1723.

79. Keizers, P. H. J.; Desreux, J. F.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2007,** *129*, 9292-9293.

80. Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997,** *268*, 158-169.

81. Battiste, J. L.; Wagner, G. *Biochemistry* **2000,** *39*, 5355-5365.

82. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005,** *127*, 476-477.

83. Liang, B.; Bushweller, J. H.; Tamm, L. K. *J. Am. Chem. Soc.* **2006,** *128*, 4389-4397.

84. Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009,** *109*, 4108-4139.

85. Otting, G. *Annu. Rev. Biophys.* **2010,** *39*, 387-405.

86. Otting, G. *J. Biomol. NMR* **2008,** *42*, 1-9.

87. Keizers, P. H. J.; Mersinli, B.; Reinle, W. G.; Donauer, J.; Hiruma, Y.; Hannemann, F.; Overhand, M.; Bernhardt, R.; Ubbink, M. *Biochemistry* **2010,** *49*, 6846-6855.

88. Koehler, J.; Meiler, J. *Prog. Nucl. Magn. Res. Sp.* **2011,** *59*, 360-389.

89. Schmitz, C.; Vernon, R.; Otting, G.; Baker, D.; Huber, T. *J. Mol. Biol.* **2012,** *416*, 668-677.

90. Pintacuda, G.; Park, A. Y.; Keniry, M. A.; Dixon, N. E.; Otting, G. *J. Am. Chem. Soc.* **2006,** *128*, 3696-3702.

91. Schmitz, C.; Stanton-Cook, M. J.; Su, X. C.; Otting, G.; Huber, T. *J. Biomol. NMR* **2008,** *41*, 179-189.

92. Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006,** *444*, 383-386.

93. Tang, C.; Ghirlando, R.; Clore, G. M. *J. Am. Chem. Soc.* **2008,** *130*, 4048-4056.

94. Ahmad, A.; Bhattacharya, A.; McDonald, R. A.; Cordes, M.; Ellington, B.; Bertelsen, E. B.; Zuiderweg, E. R. *Proc. Natl. Acad. Sci. U. S. A* **2011,** *108*, 18966-18971.

95. Zhuang, T.; Vishnivetskiy, S. A.; Gurevich, V. V.; Sanders, C. R. *Biochemistry* **2010,** *49*, 10473-10485.

96. Iwahara, J.; Clore, G. M. *Nature* **2006,** *440*, 1227-1230.

97. Iwahara, J.; Clore, G. M. *Nature* **2006,** *440*, 1227-1230.

98. Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A.; Ubbink, M. *J. Biol. Chem.* **2005,** *280*, 35784.

99. Gabdoulline, R. R.; Wade, R. C. *J. Mol. Recognit.* **1999,** *12*, 226-234.

100. Spaar, A.; Dammer, C.; Gabdoulline, R. R.; Wade, R. C.; Helms, V. *Biophys. J.* **2006,** *90*, 1913-1924.

101. Anthis, N. J.; Doucleff, M.; Clore, G. M. *J. Am. Chem. Soc.* **2011,** *133*, 18966-18974.

102. Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007,** *449*, 1078-1082.

103. Takayama, Y.; Clore, G. M. *Proc. Natl. Acad. Sci. U. S. A* **2011,** *108*, E169-E176.

104. Takayama, Y.; Clore, G. M. *J. Biol. Chem.* **2012,** *287*, 14349-14363.

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

106. Gross, E. L.; Pearson, D. C., Jr. *Biophys. J.* **2003,** *85*, 2055-2068.

107. Gross, E. L.; Rosenberg, I. *Biophys. J.* **2006,** *90*, 366-380.

108. Gross, E. L. *Photosynth. Res.* **2007,** *94*, 411-422.

109. Northrup, S. H.; Boles, J. O.; Reynolds, J. C. *Science* **1988,** *241*, 67-70.

110. Pelletier, H.; Kraut, J. *Science* **1992,** *258*, 1748-1755.

111. Schreiber, G.; Haran, G.; Zhou, H. X. *Chem Rev* **2009,** *109*, 839-860.

112. Nocek, J. M.; Knutson, A. K.; Xiong, P.; Co, N. P.; Hoffman, B. M. *J. Am. Chem. Soc.* **2010,** *132*, 6165-6175.

113. Clore, G. M. *Protein Sci.* **2011,** *20*, 229-246.

114. Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 18945-18950.

115. Ubbink, M.; Ejdebäck, M.; Karlsson, B. G.; Bendall, D. S. *Structure* **1998,** *6*, 323-335.

116. Lange, C.; Cornvik, T.; Diaz-Moreno, I.; Ubbink, M. *Biochim. Biophys. Acta-Bioenerg.* **2005,** *1707*, 179-188.

117. Sato, K.; Kohzuma, T.; Dennison, C. *J. Am. Chem. Soc.* **2003,** *125*, 2101-2112.

118. Schlarb-Ridley, B. G.; Bendall, D. S.; Howe, C. J. *Biochemistry* **2002,** *41*, 3279-3285.

119. Hart, S. E.; Schlarb-Ridley, B. G.; Delon, C.; Bendall, D. S.; Howe, C. J. *Biochemistry* **2003,** *42*, 4829-4836.

120. Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A.; Ubbink, M. *J. Biol. Chem.* **2005,** *280*, 18908-18915.

121. Albarran, C.; Navarro, J. A.; Molina-Heredia, F. P.; Murdoch, P. S.; De la Rosa, M. A.; Hervas, M. *Biochemistry* **2005,** *44*, 11601-11607.

122. Albarran, C.; Navarro, J. A.; De la Rosa, M. A.; Hervas, M. *Biochemistry* **2007,** *46*, 997-1003.

123. Milikisyants, S.; Scarpelli, F.; Finiguerra, M. G.; Ubbink, M.; Huber, M. *J. Magn. Reson.* **2009,** *201*, 48-56.

124. Arslan, E.; Schulz, H.; Zufferey, R.; Kunzler, P.; Thony-Meyer, L. *Biochem. Biophys. Res. Commun.* **1998,** *251*, 744-747.

125. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995,** *6*, 277-293.

126. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins* **2005,** *59*, 687-696.

127. Kannt, A.; Young, S.; Bendall, D. S. *Biochim. Biophys. Acta-Bioenerg.* **1996,** *1277*, 115-126.

128. Battiste, J. L.; Wagner, G. *Biochemistry* **2000,** *39*, 5355-5365.

129. de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *J. Magn. Reson.* **2000,** *147*, 138-146.

130. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003,** *160*, 65-73.

131. Baniulis, D.; Yamashita, E.; Whitelegge, J. P.; Zatsman, A. I.; Hendrich, M. P.; Hasan, S. S.; Ryan, C. M.; Cramer, W. A. *J. Biol. Chem.* **2009,** *284*, 9861-9869.

132. Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004,** *126*, 5879-5896.

133. Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A.; Ubbink, M. *J. Biol. Chem.* **2005,** *280*, 18908-18915.

134. Ubbink, M.; Lian, L. Y.; Modi, S.; Evans, P. A.; Bendall, D. S. *Eur. J. Biochem.* **1996,** *242*, 132-147.

135. Diaz-Moreno, I.; Diaz-Quintana, A.; De la Rosa, M. A.; Crowley, P. B.; Ubbink, M. *Biochemistry* **2005,** *44*, 3176-3183.

136. Fawzi, N. L.; Doucleff, M.; Suh, J. Y.; Clore, G. M. *Proc. Natl. Acad. Sci. U. S. A* **2010,** *107*, 1379-1384.

137. Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997,** *72*, 1917-1929.

138. Camacho, C. J.; Weng, Z.; Vajda, S.; DeLisi, C. *Biophys. J.* **1999,** *76*, 1166-1178.

139. Camacho, C. J.; Vajda, S. *Proc. Natl. Acad. Sci. U. S. A* **2001,** *98*, 10636-10641.

140. Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007,** *447*, 1021-1025.

141. Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006,** *444*, 383-386.

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

143. Milikisyants, S.; Scarpelli, F.; Finiguerra, M. G.; Ubbink, M.; Huber, M. *J. Magn. Reson.* **2009,** *201*, 48-56.

144. Arslan, E.; Schulz, H.; Zufferey, R.; Kunzler, P.; Thony-Meyer, L. *Biochem. Biophys. Res. Commun.* **1998,** *251*, 744-747.

145. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995,** *6*, 277-293.

146. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins* **2005,** *59*, 687-696.

147. Baniulis, D.; Yamashita, E.; Whitelegge, J. P.; Zatsman, A. I.; Hendrich, M. P.; Hasan, S. S.; Ryan, C. M.; Cramer, W. A. *J. Biol. Chem.* **2009,** *284*, 9861-9869.

148. Brunger, A. T.; Karplus, M. *Proteins* **1988,** *4*, 148-156.

149. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983,** *4*, 187-217.

150. MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998,** *102*, 3586-3616.

151. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A* **2001,** *98*, 10037-10041.

152. Metropolis, N.; Ulam S. *J. Am. Stat. Assoc.* **1949,** *44*, 335-341.

153. García de la Torre, J.; Huertas, M. L.; Carrasco, B. *J. Magn. Reson.* **2000,** *147*, 138-146.

154. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn Reson.* **2003,** *160*, 65-73.

155. Ubbink, M.; Lian, L. Y.; Modi, S.; Evans, P. A.; Bendall, D. S. *Eur. J. Biochem.* **1996,** *242*, 132-147.

156. Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007,** *449*, 1078-1082.

157. Tang, C.; Ghirlando, R.; Clore, G. M. *J. Am. Chem. Soc.* **2008,** *130*, 4048-4056.

158. Tang, C.; Louis, J. M.; Aniana, A.; Suh, J. Y.; Clore, G. M. *Nature* **2008,** *455*, 693-696.

159. Fawzi, N. L.; Doucleff, M.; Suh, J. Y.; Clore, G. M. *Proc. Natl. Acad. Sci. U. S. A* **2010,** *107*, 1379-1384.

160. Anthis, N. J.; Doucleff, M.; Clore, G. M. *J. Am. Chem. Soc.* **2011,** *133*, 18966-18974.

161. Volkov, A. N.; Ubbink, M.; van Nuland, N. A. J. *J. Biomol. NMR* **2010,** *48*, 225-236.

162. Moser, C. C.; Keske, J. M.; Warncke, K.; Farid, R. S.; Dutton, P. L. *Nature* **1992,** *355*, 796-802.

163. Gabdoulline, R. R.; Wade, R. C. *J. Mol. Biol.* **2001,** *306*, 1139-1155.

164. Hasan, S. S.; Cramer, W. A. *Phys. Chem. Chem. Phys.* **2012,** *14*, 13853-13860.

165. Harel, M.; Cohen, M.; Schreiber, G. *J. Mol. Biol.* **2007,** *371*, 180-196.

166. Liang, Z. X.; Nocek, J. M.; Huang, K.; Hayes, R. T.; Kurnikov, I. V.; Beratan, D. N.; Hoffman, B. M. *J. Am. Chem. Soc.* **2002,** *124*, 6849-6859.

167. Hope, A. B. *Biochim. Biophys. Acta* **2000,** *1456*, 5-26.

168. Diaz-Quintana, A.; Hervás, M.; Navarro, J.A.; De la Rosa, M. A. *Photosynthetic protein complexes: A structural approach.* (Fromme P. Ed.) Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. **2008,** 181-200.

169. Camacho, C. J.; Weng, Z.; Vajda, S.; DeLisi, C. *Biophys J* **1999,** *76*, 1166-1178.

170. Camacho, C. J.; Kimura, S. R.; DeLisi, C.; Vajda, S. *Biophys. J.* **2000,** *78*, 1094-1105.

171. Kim, Y. C.; Tang, C.; Clore, G. M.; Hummer, G. *Proc. Natl. Acad. Sci. U. S. A* **2008,** *105*, 12855-12860.

172. Musiani, F.; Dikiy, A.; Semenov, A.Y.; Ciurli, S. *J. Biol. Chem.* **2005,** *280*, 18833-18841.

173. Schlarb, B. G.; Wagner, M. J.; Vijgenboom, E.; Ubbink, M.; Bendall, D. S.; Howe, C. J. *Gene* **1999,** *234*, 275-283.

174. Banci, L.; Bertini, I.; Cavallaro, G.; Giachetti, A.; Luchinat, C.; Parigi, G. *J. Biomol. NMR* **2004,** *28*, 249-261.

175. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn Reson.* **2003,** *160*, 65-73.

176. Kasha, M. *J. Chem. Phys.* **1952,** *20*, 71-74.

177. Marcus, R. A. *J. Chem. Phys.* **1956,** *24*, 966-978.

178. Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985,** *811*, 265-322.

179. Northrup, S. H.; Boles, J. O.; Reynolds, J. C. L. *J. Phys. Chem.* **1987,** *91*, 5991-5998.

180. Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *Chembiochem.* **2005,** *6*, 1536-1549.

181. Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009,** *109*, 4108-4139.

182. Clore, G. M.; Schwieters, C. D. *Curr. Opin. Struct. Biol.* **2002,** *12*, 146-153.

183. Gabdoulline, R. R.; Wade, R. C. *Methods* **1998,** *14*, 329-341.

184. Gabdoulline, R. R.; Wade, R. C. *Curr. Opin. Struct. Biol.* **2002,** *12*, 204-213.

185. Smith, G. R.; Sternberg, M. J. *Curr. Opin. Struct. Biol.* **2002,** *12*, 28-35.

186. Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. *J. Am. Chem. Soc.* **2003,** *125*, 1731-1737.

187. Wider, G. *Nuclear Magnetic Resonance of Biological Macromolecules, Part C* **2005,** *394*, 382-398.

188. Kay, L. E. *J. Magn Reson.* **2011,** *210*, 159-170.

189. Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J.; Vajda, S.; Vakser, I.; Wodak, S. J. *Proteins: Struct. Funct. Genet.* **2003,** *52*, 2-9.

190. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. *Bioinformatics.* **2004,** *20*, 45-50.

Appendices

Appendix 1

Input file for rigid-body docking of Cyt *f*-Pc using PRE restraints for Xplor-NIH 2.9.9

```
parameter
        @nostoc_cytf_SL_pc.par
 end
 structure
        @nostoc_cytf_SL_pc.psf
 end
 eval ($inifile="nostoc_cytf_SL_pc.pdb")
 coordinates @$inifile
 @learn.pcf.par
 constraints fix (segid="CYTF" or resn SL) end
 eval ($a05 = 6000)                    !total nr cycles / $a53
 eval ($a53 = 1000)                    !nr of steps
 eval ($u2 = 30)              !nr of cycles per docking approach
 eval ($a54 = 0.01)           !timestep in ps
 eval ($a61 = $a54*$a53)      !time per cycle (ps)
 eval ($a80 = $a61*$a05)      !total time (ps)
 eval ($a58=30.0)             !velocity factor
 eval ($a56=1.5)              !fbeta
 eval ($a55=300.0)            !TBATH
 eval ($a59=45.0)             !.pdb writing threshold
 eval ($van=100.0)            ! vdw writing threshold
 eval ($a59a=0.01*$a59)                !vx-reset minimum
 eval ($a60=1.0)              !.pdb lag factor              !5
 eval ($a301=1.0)             !VDW repel scale factor
 eval ($a93=1)                !time-spent factor: 0<$a93<1 (fraction bound)
 eval ($g1a = "N71C")         !names of the spin label 1 used in this calculation
 eval ($g1b = "Q104C")        !names of the spin label 2 used in this calculation
 eval ($g1c = "S192C")        !names of the spin label 3 used in this calculation
 eval ($g2a = 13)    !residue number of the 1st SL conformation of SL 1
 eval ($g2b = 17)    !residue number of the 1st SL conformation of SL 2
 eval ($g2c = 21)    !residue number of the 1st SL conformation of SL 3
 eval ($field=600.1328E6)     !field strength in Hz
 eval ($frl=1)                         !fraction of proteins with spin label
 eval ($tau_c=30.0E-9)                 !tau c for the complex in sec
 eval ($a9=0.01)                       !general scaling PRE (NOE) term
 eval ($a10=1.0)              !scale factor for CL1 (peaks disappeared)
 eval ($a11=1.0)              !scale factor for CL2 (peaks unaffected)
 eval ($a12=1.0)              !scale factor for CL3 (peaks reduced)
 eval ($low1=14.0)   !lower(d_minus) limit (SLA restraints upper limit only)
 eval ($up1=4.0)    !upper(d_plus) limit (restraints with upper limit only)
 eval ($low2=4.0)    !lower(d_minus) limit (restraints with lower limit only)
 eval ($up2=100.0)   !upper(d_plus) limit (restraints with lower limit only)
 eval ($low3=4.0)    !lower(d_minus) limit (restraints with both limits)
 eval ($up3=4.0)     !upper(d_plus) limit (restraints with both limits)
 eval ($a01 = 1)              !cycle counter
 eval ($a48=$cpu*1e4)
 set seed=$a48 end
 eval ($a14=0)
 eval ($a18=1)
 eval ($nout1=0)     ! Number of strucutres output per run
 eval ($min1=9999.0)
 eval ($min2=9999.0)
 eval ($ref=0)
 set display=coor.dat end              !write parameters
 display xx ----------PARAMETERS----------
 display xx startdate:       $DATE
 display xx starttime:       $TIME
 display xx ini.file:        $inifile
 display xx timestep (ps):   $a54
```

```
 display xx nr of steps:       $a53              time /cycle (ps):   $a61
 display xx nr of cycles:      $a05              total time  (ps):   $a80
 display xx time-spent fact:   $a93
 display xx general scale:     $a9
 display xx TBATH:             $a55    velocity factor:   $a56   fbeta:   $a58
 display xx .pdb threshold     $a59
 display xx .pdb lag factor    $a60
 display xx Local mimim. impulse after 10 cycles with 'constant' Etot > $a59a
 display xx ----------------------------------------------------------
 display
 set display=OUTPUT end
 flag exclude elec bond angl dihe impr include vdw noe end
 vector do (fbeta=$a58) (segid="PC")
 vector do (vx=$a56) (segid="PC")
 vector do (vy=$a56) (segid="PC")
 vector do (vz=$a56) (segid="PC")
set disp=ener.dat end
display Energies for the output structures
display file                Etot          VDW         NOE
display ------------------------------------------------------
set disp=OUTPUT end
@restraints.xpl
eval ($a213=100*(rand()-0.5) ) ! push away PC
eval ($a214=100*(rand()-0.5) )
eval ($a215=100*(rand()-0.5) )
vector do (x=x+$a213) (segid="PC")
vector do (y=y+$a214) (segid="PC")
vector do (z=z+$a215) (segid="PC")
parameter
     @nbfix.4sl_expl.xpl
      nbonds
            cutnb=8.5
            inhi=0.25
            ctofnb=7.5
            ctonnb=6.5
            repe1=0.6 !0.6
                        NBXMod=-2
            rexp=2
            irex=2
            rcon=$a301
            wmin=1.5
       end
 end
 constraints
              interactions (segid="CYTF")(segid="PC")
 end
 energy end
 while ($a01 LE $a05) loop calc
  display cycle $a01
  dynamics rigid
              dt=$a54
              group=(segid="PC")
              dynmode=TCOU
              tbath=$a55
              nprint=500
              nstep=$a53
              NTRFRQ=0 !new for XPLOR vs 3.8
  end
 eval ($a14=$ENER)
 if ($a14 < $min2) then
   coor copy end
   eval ($min2 = $a14)
   eval ($abc=$VDW)
   eval ($a15=$NOE)
```

```
  eval ($a16=$a01)
  eval ($a17=$a18)
end if
if ($a18 = $u2) then                !counting number of dockings
 if ($min2 < $a59) then
  if ($abc < $van) then
            coor swap end
            eval ($nout1=$nout1+1)
            eval ($pdb="structure_"+encode($nout1)+".pdb")
            write coord output=$pdb end
            if ($a14<$min1) then
             eval ($ref=$nout1)
       eval ($min1 = $a14)
      end if
      pick bond (segid="CYTF" and name FE) (segid="PC" and name CU) geom
      eval ($a50=$RESULT)
      set disp=ener.dat end
      display $nout1  $min2       $abc        $a15
      set disp=OUTPUT end
      set disp=coor.dat end
      display Cycle: $a16         Dock step: $a17      File#: $nout1           Fe-Cu: $a50             temp: $TEMP
      display  Etot:    $min2                    Evdw:  $abc              Enoe:  $a15
      display ------------------------------------------------------------
            set disp=OUTPUT end
    !violation analysis
            eval ($b3="structure_"+encode($nout1)+".viol.dat")
            set display=$b3 end
            display Violation analysis PRE data
            display res#   PRE(meas)  PRE(calc)  Dist(calc)
            display
            for $g2 in ($g2a $g2b $g2c)
            loop spinlabel
              eval ($g2d=$g2)
              eval ($g2e=$g2+1)
              eval ($g2f=$g2+2)
              eval ($g2g=$g2+3)
              if ($g2=$g2a) then
                eval ($g1h = $g1a)
                display Spin Label $g1h
                for
                        $c02 in ID (store1)
                loop C3
                  vector show elem (resi) (ID $c02)
                  eval ($c04=$RESULT)                                   !residue nr
                  vector show elem (store1) (ID $c02)
                  eval ($c08=$RESULT)
                  pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
                  eval ($ca1=$RESULT)
                  pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
                  eval ($ca2=$RESULT)
                  pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
                  eval ($ca3=$RESULT)
                  pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
                  eval ($ca4=$RESULT)
                  eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                                !r-6 average distance over 4 positions
                  eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                                !back calculated PRE
                  eval ($c10=$c03^(-1/6))
                  display $c04       $c08        $c05         $c10
                end loop C3
                display
              end if
              if ($g2=$g2b) then
```

```
                eval ($g1h = $g1b)
                display Spin Label $g1h
                for
                        $c02 in ID (store2)
                loop C3
                  vector show elem (resi) (ID $c02)
                  eval ($c04=$RESULT)                                   !residue nr
                  vector show elem (store2) (ID $c02)
                  eval ($c08=$RESULT)
                  pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
                  eval ($ca1=$RESULT)
                  pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
                  eval ($ca2=$RESULT)
                  pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
                  eval ($ca3=$RESULT)
                  pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
                  eval ($ca4=$RESULT)
                  eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                                !r-6 average distance over 4 positions
                  eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                                !back calculated PRE
                 eval ($c10=$c03^(-1/6))
                  display $c04      $c08       $c05        $c10
                end loop C3
                display
               end if
               if ($g2=$g2c) then
                eval ($g1h = $g1c)
                display Spin Label $g1h
                for
                        $c02 in ID (store3)
                loop C3
                  vector show elem (resi) (ID $c02)
                  eval ($c04=$RESULT)                                   !residue nr
                  vector show elem (store3) (ID $c02)
                  eval ($c08=$RESULT)
                  pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
                  eval ($ca1=$RESULT)
                  pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
                  eval ($ca2=$RESULT)
                  pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
                  eval ($ca3=$RESULT)
                  pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
                  eval ($ca4=$RESULT)
                  eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                                !r-6 average distance over 4 positions
                  eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                                !back calculated PRE
                  eval ($c10=$c03^(-1/6))
                  display $c04      $c08       $c05        $c10
                end loop C3
                display
               end if
      end loop spinlabel
             close $b3 end
             set display=OUTPUT end
    !end violoation analysis
     coor swap end
    end if
  end if
!impulse to escape local minimum
  eval ($a213=100*(rand()-0.5) )
  eval ($a214=100*(rand()-0.5) )
  eval ($a215=100*(rand()-0.5) )
```

```
  vector do (x=x+$a213) (segid="PC")
  vector do (y=y+$a214) (segid="PC")
  vector do (z=z+$a215) (segid="PC")
  vector do (vx=50) (segid="PC")
  vector do (vy=50) (segid="PC")
  vector do (vz=50) (segid="PC")
  set disp=coor.dat end
  display cycle $a01: impulse: $a213, $a214, $a215 to x,y,z; v=50
  set disp=OUTPUT end
 eval ($a18=0)
 eval ($min2=9999.0)
 end if
 eval ($a18=$a18+1)
  if ($TEMP>200000.0) then                    !correct excessive temperature
    eval ($a45=RAND()+0.01)
    vector do (vx=$a56*$a45*rand()) (segid="PC")
    vector do (vy=$a56*$a45*rand()) (segid="PC")
    vector do (vz=$a56*$a45*rand()) (segid="PC")
   end if
   eval ($a01 = $a01 + 1)
 end loop calc
if ($ref > 0) then
  vector idend ( store9 ) ( name ca or name n or name c )        !backbone selection
  eval ($ref_file="structure_"+encode($ref)+".pdb")
  set display=rms.dat end
  display Backbone pairwise RMSD from the lowest energy strucutre ($ref_file)
  display  file          rmsd
  display
  set disp=OUTPUT end
    coor swap end
    coor init end
    coor swap end
    coor disp=comp @@$ref_file
  eval ($count1=0)
  while ($count1 < $nout1) loop fill
   evaluate ($count1=$count1+1)
        evaluate ($file="structure_"+encode($count1)+".pdb")
        coor init end
        coor @@$file
        coor sele=(recall 9) fit end
        coor sele=(recall 9) rms end
        eval ($b1 = $result)
   set display=rms.dat end
   display  $file          $b1
   end loop fill
end if
set display=OUTPUT end
 set echo=true end
  stop
```

Appendix 2

Input file to determine distances from Cyt *f*-SL to Pc amide protons for Xplor-NIH 2.9.9

```
parameter
        @nostoc_cytf_SL_pc.par
end
structure
        @nostoc_cytf_SL_pc.psf
end
eval ($inifile="nostoc_cytf_SL_pc.pdb")
coordinates @$inifile
@learn.pcf.par
constraints fix (segid="CYTF" or resn SL) end
eval ($a05 = 6000)                      !total nr cycles / $a53
eval ($a53 = 1000)                      !nr of steps
eval ($u2 = 30)                 !nr of cycles per docking approach
eval ($a54 = 0.01)              !timestep in ps
eval ($a61 = $a54*$a53)         !time per cycle (ps)
eval ($a80 = $a61*$a05)         !total time (ps)
eval ($a58=30.0)                !velocity factor
eval ($a56=1.5)                 !fbeta
eval ($a55=300.0)               !TBATH
eval ($a59=45.0)                !.pdb writing threshold
eval ($van=100.0)               ! vdw writing threshold
eval ($a59a=0.01*$a59)                  !vx-reset minimum
eval ($a60=1.0)                 !.pdb lag factor
eval ($a301=1.0)                !VDW repel scale factor
eval ($a93=1)                   !time-spent factor: 0<$a93<1 (fraction bound)
eval ($g1a = "N71C") !names of the spin label 1 used in calculation
eval ($g1b = "Q104C")           !names of the spin label 2 used in calculation
eval ($g1c = "S192C") !names of the spin label 3 used in calculation
eval ($g2a = 13)    !residue number of the 1st SL conformation of SL 1
eval ($g2b = 17)    !residue number of the 1st SL conformation of SL 2
eval ($g2c = 21)    !residue number of the 1st SL conformation of SL 3
eval ($field=600.1328E6)        !field strength in Hz
eval ($frl=1)                           !fraction of proteins with spin label
eval ($tau_c=30.0E-9)                   !tau c for the complex in sec
eval ($a9=0.01)                         !general scaling PRE (NOE) term
eval ($a10=1.0)     !scale factor for CL1 (peaks disappeared)
eval ($a11=1.0)     !scale factor for CL2 (peaks unaffected)
eval ($a12=1.0)                 !scale factor for CL3 (peaks reduced)
eval ($low1=14.0)   !lower(d_minus) limit
eval ($up1=4.0)     !upper(d_plus) limit
eval ($low2=4.0)                !lower(d_minus) limit
eval ($up2=100.0)                       !upper(d_plus) limit
eval ($low3=4.0)                !lower(d_minus) limit
eval ($up3=4.0)                 !upper(d_plus) limit
eval ($a01 = 1)                 !cycle counter
eval ($a48=$cpu*1e4)
set seed=$a48 end
eval ($a14=0)
eval ($a18=1)
eval ($nout1=0)     ! Number of strucutres output per run
eval ($min1=9999.0)
eval ($min2=9999.0)
eval ($ref=0)
set display=coor.dat end                !write parameters
display xx ----------PARAMETERS----------
display xx startdate:       $DATE
```

```
 display xx starttime:        $TIME
 display xx ini.file:          $inifile
 display xx timestep (ps):    $a54
 display xx nr of steps:      $a53              time /cycle (ps):   $a61
 display xx nr of cycles:     $a05              total time  (ps):   $a80
 display xx time-spent fact:  $a93
 display xx general scale:    $a9
 display xx TBATH:            $a55  velocity factor: $a56 fbeta:   $a58
 display xx .pdb threshold    $a59
 display xx .pdb lag factor   $a60
 display xx Local mimim. impulse after 10 cycles with 'constant' Etot > $a59a
 display xx ------------------------------------------------------------
 display
 set display=OUTPUT end
 flag exclude elec bond angl dihe impr include vdw noe end
 vector do (fbeta=$a58) (segid="PC")
 vector do (vx=$a56) (segid="PC")
 vector do (vy=$a56) (segid="PC")
 vector do (vz=$a56) (segid="PC")
set disp=ener.dat end
display Energies for the output structures
display file              Etot          VDW         NOE
display ----------------------------------------------------
set disp=OUTPUT end
@restraints.xpl
eval ($a213=100*(rand()-0.5) ) ! push away PC
eval ($a214=100*(rand()-0.5) )
eval ($a215=100*(rand()-0.5) )
vector do (x=x+$a213) (segid="PC")
vector do (y=y+$a214) (segid="PC")
vector do (z=z+$a215) (segid="PC")
!write coor OUTPUT="test1.pdb" end
parameter
     @nbfix.4sl_expl.xpl
      nbonds
            cutnb=8.5
            inhi=0.25
            ctofnb=7.5
            ctonnb=6.5
            repe1=0.6 !0.6
                        NBXMod=-2
            rexp=2
            irex=2
            rcon=$a301
            wmin=1.5
      end
 end
 constraints
             interactions (segid="CYTF")(segid="PC")
 end
 energy end
 while ($a01 LE $a05) loop calc
  display cycle $a01
  dynamics rigid
             dt=$a54
             group=(segid="PC")
             dynmode=TCOU
             tbath=$a55
             nprint=500
             nstep=$a53
             NTRFRQ=0 !new for XPLOR vs 3.8
  end
 eval ($a14=$ENER)
 if ($a14 < $min2) then
```

```
  coor copy end
  eval ($min2 = $a14)
  eval ($abc=$VDW)
  eval ($a15=$NOE)
  eval ($a16=$a01)
  eval ($a17=$a18)
 end if
 if ($a18 = $u2) then                !counting number of dockings
  if ($min2 < $a59) then
   if ($abc < $van) then
              coor swap end
              eval ($nout1=$nout1+1)
              eval ($pdb="structure_"+encode($nout1)+".pdb")
              write coord output=$pdb end
              if ($a14<$min1) then
               eval ($ref=$nout1)
        eval ($min1 = $a14)
       end if
       pick bond (segid="CYTF" and name FE) (segid="PC" and name CU) geom
       eval ($a50=$RESULT)
       set disp=ener.dat end
       display $nout1  $min2       $abc        $a15
       set disp=OUTPUT end
       set disp=coor.dat end
       display Cycle: $a16          Dock step: $a17       File#: $nout1            Fe-Cu: $a50              temp: $TEMP
       display  Etot:    $min2                      Evdw:  $abc              Enoe:  $a15
       display ------------------------------------------------------------
              set disp=OUTPUT end
     !violation analysis
              eval ($b3="structure_"+encode($nout1)+".viol.dat")
              set display=$b3 end
              display Violation analysis PRE data
              display res#   PRE(meas)  PRE(calc)  Dist(calc)
              display
              for $g2 in ($g2a $g2b $g2c)
              loop spinlabel
               eval ($g2d=$g2)
               eval ($g2e=$g2+1)
               eval ($g2f=$g2+2)
               eval ($g2g=$g2+3)
               if ($g2=$g2a) then
                eval ($g1h = $g1a)
                display Spin Label $g1h
                for
                          $c02 in ID (store1)
                loop C3
                  vector show elem (resi) (ID $c02)
                  eval ($c04=$RESULT)                                  !residue nr
                  vector show elem (store1) (ID $c02)
                  eval ($c08=$RESULT)
                  pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
                  eval ($ca1=$RESULT)
                  pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
                  eval ($ca2=$RESULT)
                  pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
                  eval ($ca3=$RESULT)
                  pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
                  eval ($ca4=$RESULT)
                  eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                                  !r-6 average distance over 4 positions
                  eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                                  !back calculated PRE
                  eval ($c10=$c03^(-1/6))
                  display $c04       $c08        $c05         $c10
```

```
      end loop C3
      display
     end if

     if ($g2=$g2b) then
      eval ($g1h = $g1b)
      display Spin Label $g1h
      for
            $c02 in ID (store2)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store2) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
     if ($g2=$g2c) then
      eval ($g1h = $g1c)
      display Spin Label $g1h
      for
            $c02 in ID (store3)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store3) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
   end loop spinlabel
        close $b3 end
        set display=OUTPUT end
 !end violoation analysis
  coor swap end
end if
```

```
  end if
                                                !impulse to escape local minimum
  eval ($a213=100*(rand()-0.5) )
  eval ($a214=100*(rand()-0.5) )
  eval ($a215=100*(rand()-0.5) )
  vector do (x=x+$a213) (segid="PC")
  vector do (y=y+$a214) (segid="PC")
  vector do (z=z+$a215) (segid="PC")
  vector do (vx=50) (segid="PC")
  vector do (vy=50) (segid="PC")
  vector do (vz=50) (segid="PC")
  set disp=coor.dat end
  display cycle $a01: impulse: $a213, $a214, $a215 to x,y,z; v=50
  set disp=OUTPUT end
 eval ($a18=0)
 eval ($min2=9999.0)
 end if
 eval ($a18=$a18+1)
  if ($TEMP>200000.0) then                    !correct excessive temperature
    eval ($a45=RAND()+0.01)
    vector do (vx=$a56*$a45*rand()) (segid="PC")
    vector do (vy=$a56*$a45*rand()) (segid="PC")
    vector do (vz=$a56*$a45*rand()) (segid="PC")
   end if
   eval ($a01 = $a01 + 1)
 end loop calc
if ($ref > 0) then
  vector idend ( store9 ) ( name ca or name n or name c )        !backbone selection
  eval ($ref_file="structure_"+encode($ref)+".pdb")
  set display=rms.dat end
  display Backbone pairwise RMSD from the lowest energy strucutre ($ref_file)
  display  file            rmsd
  display ----------------------------------------------------------------
  set disp=OUTPUT end
    coor swap end
    coor init end
    coor swap end
    coor disp=comp @@$ref_file
  eval ($count1=0)
  while ($count1 < $nout1) loop fill
   evaluate ($count1=$count1+1)
        evaluate ($file="structure_"+encode($count1)+".pdb")
        coor init end
        coor @@$file
        coor sele=(recall 9) fit end
        coor sele=(recall 9) rms end
        eval ($b1 = $result)
   set display=rms.dat end
   display  $file           $b1
   end loop fill
end if
set display=OUTPUT end
 set echo=true end
  stop
```

Appendix 3

Input file for ensemble docking of Cyt *f*-Pc using PRE restraints for Xplor-NIH 2.9.9

```
parameter
        @nostoc_cytf_SL_pc.par
end
structure
        @nostoc_cytf_SL_pc.psf
end
eval ($inifile="nostoc_cytf_SL_pc.pdb")
coordinates @$inifile
@learn.pcf.par
constraints fix (segid="CYTF" or resn SL) end
eval ($a05 = 6000)                      !total nr cycles / $a53
eval ($a53 = 1000)                      !nr of steps
eval ($u2 = 30)                 !nr of cycles per docking approach
eval ($a54 = 0.01)              !timestep in ps
eval ($a61 = $a54*$a53)         !time per cycle (ps)
eval ($a80 = $a61*$a05)         !total time (ps)
eval ($a58=30.0)                !velocity factor
eval ($a56=1.5)                 !fbeta
eval ($a55=300.0)               !TBATH
eval ($a59=45.0)                !.pdb writing threshold
eval ($van=100.0)               ! vdw writing threshold
eval ($a59a=0.01*$a59)                  !vx-reset minimum
eval ($a60=1.0)                 !.pdb lag factor
eval ($a301=1.0)                !VDW repel scale factor
eval ($a93=1)                   !time-spent factor: 0<$a93<1 (fraction bound)
eval ($g1a = "N71C")            !names of the spin label 1
eval ($g1b = "Q104C")                   !names of the spin label 2
eval ($g1c = "S192C")                   !names of the spin label 3
eval ($g2a = 13)     !residue number of the 1stSL conformation of SL 1
eval ($g2b = 17)     !residue number of the 1st SL conformation of SL 2
eval ($g2c = 21)     !residue number of the 1st SL conformation of SL 3
eval ($field=600.1328E6)        !field strength in Hz
eval ($frl=1)                           !fraction of proteins with spin label
eval ($tau_c=30.0E-9)                   !tau c for the complex in sec
eval ($a9=0.01)                         !general scaling PRE (NOE) term
eval ($a10=1.0)      !scale factor for CL1 (peaks disappeared)
eval ($a11=1.0)                 !scale factor for CL2 (peaks unaffected)
eval ($a12=1.0)                 !scale factor for CL3 (peaks reduced)
eval ($low1=14.0)  !lower(d_minus) limit (restraints upper limit only)
eval ($up1=4.0)    !upper(d_plus) limit (restraints upper limit only)
eval ($low2=4.0)     !lower(d_minus) limit (restraints lower limit only)
eval ($up2=100.0)   !upper(d_plus) limit (restraints lower limit only)
eval ($low3=4.0)   !lower(d_minus) limit (restraints with both limits)
eval ($up3=4.0)    !upper(d_plus) limit (restraints with both limits)
eval ($a01 = 1)                 !cycle counter
eval ($a48=$cpu*1e4)
set seed=$a48 end
eval ($a14=0)
eval ($a18=1)
eval ($nout1=0)      ! Number of strucutres output per run
eval ($min1=9999.0)
eval ($min2=9999.0)
eval ($ref=0)
set display=coor.dat end                !write parameters
display xx ----------PARAMETERS----------
display xx startdate:       $DATE
display xx starttime:       $TIME
```

```
 display xx ini.file:            $inifile
 display xx timestep (ps):     $a54
 display xx nr of steps:       $a53              time /cycle (ps):   $a61
 display xx nr of cycles:      $a05              total time  (ps):   $a80
 display xx time-spent fact:   $a93
 display xx general scale:     $a9
 display xx TBATH:      $a55    velocity factor:   $a56   fbeta:   $a58
 display xx .pdb threshold     $a59
 display xx .pdb lag factor    $a60
 display xx Local mimim. impulse after 10 cycles with 'constant' Etot > $a59a
 display xx ------------------------------------------------------------
 display
 set display=OUTPUT end
 flag exclude elec bond angl dihe impr include vdw noe end
 vector do (fbeta=$a58) (segid="PC")
 vector do (vx=$a56) (segid="PC")
 vector do (vy=$a56) (segid="PC")
 vector do (vz=$a56) (segid="PC")
set disp=ener.dat end
display Energies for the output structures
display file                Etot          VDW         NOE
display ------------------------------------------------
set disp=OUTPUT end
@restraints.xpl
eval ($a213=100*(rand()-0.5) ) ! push away PC
eval ($a214=100*(rand()-0.5) )
eval ($a215=100*(rand()-0.5) )
vector do (x=x+$a213) (segid="PC")
vector do (y=y+$a214) (segid="PC")
vector do (z=z+$a215) (segid="PC")
!write coor OUTPUT="test1.pdb" end
parameter
     @nbfix.4sl_expl.xpl
      nbonds
            cutnb=8.5
            inhi=0.25
            ctofnb=7.5
            ctonnb=6.5
            repe1=0.6 !0.6
                        NBXMod=-2
            rexp=2
            irex=2
            rcon=$a301
            wmin=1.5
       end
 end
 constraints
              interactions (segid="CYTF")(segid="PC")
 end
 energy end
 while ($a01 LE $a05) loop calc
  display cycle $a01
  dynamics rigid
               dt=$a54
               group=(segid="PC")
               dynmode=TCOU
               tbath=$a55
               nprint=500
               nstep=$a53
               NTRFRQ=0 !new for XPLOR vs 3.8
  end
 eval ($a14=$ENER)
 if ($a14 < $min2) then
   coor copy end
```

```
  eval ($min2 = $a14)
  eval ($abc=$VDW)
  eval ($a15=$NOE)
  eval ($a16=$a01)
  eval ($a17=$a18)
end if
if ($a18 = $u2) then                !counting number of dockings
 if ($min2 < $a59) then
  if ($abc < $van) then
            coor swap end
            eval ($nout1=$nout1+1)
            eval ($pdb="structure_"+encode($nout1)+".pdb")
            write coord output=$pdb end
            if ($a14<$min1) then
             eval ($ref=$nout1)
       eval ($min1 = $a14)
      end if
      pick bond (segid="CYTF" and name FE) (segid="PC" and name CU) geom
      eval ($a50=$RESULT)
      set disp=ener.dat end
      display $nout1  $min2      $abc       $a15
      set disp=OUTPUT end
      set disp=coor.dat end
      display Cycle: $a16         Dock step: $a17      File#: $nout1          Fe-Cu: $a50             temp: $TEMP
      display  Etot:    $min2                  Evdw:  $abc             Enoe:  $a15
      display ------------------------------------------------------------
            set disp=OUTPUT end
    !violation analysis
            eval ($b3="structure_"+encode($nout1)+".viol.dat")
            set display=$b3 end
            display Violation analysis PRE data
            display res#   PRE(meas)  PRE(calc)  Dist(calc)
            display
            for $g2 in ($g2a $g2b $g2c)
            loop spinlabel
              eval ($g2d=$g2)
              eval ($g2e=$g2+1)
              eval ($g2f=$g2+2)
              eval ($g2g=$g2+3)
              if ($g2=$g2a) then
                eval ($g1h = $g1a)
                display Spin Label $g1h
                for
                        $c02 in ID (store1)
                loop C3
                  vector show elem (resi) (ID $c02)
                  eval ($c04=$RESULT)                                !residue nr
                  vector show elem (store1) (ID $c02)
                  eval ($c08=$RESULT)
                  pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
                  eval ($ca1=$RESULT)
                  pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
                  eval ($ca2=$RESULT)
                  pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
                  eval ($ca3=$RESULT)
                  pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
                  eval ($ca4=$RESULT)
                  eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                                !r-6 average distance over 4 positions
                  eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                                !back calculated PRE
                  eval ($c10=$c03^(-1/6))
                  display $c04       $c08        $c05         $c10
                end loop C3
```

```
      display
     end if
     if ($g2=$g2b) then
      eval ($g1h = $g1b)
      display Spin Label $g1h
      for
            $c02 in ID (store2)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store2) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
     if ($g2=$g2c) then
      eval ($g1h = $g1c)
      display Spin Label $g1h
      for
            $c02 in ID (store3)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store3) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
  end loop spinlabel
      close $b3 end
      set display=OUTPUT end
 !end violoation analysis
  coor swap end
 end if
end if
!impulse to escape local minimum
```

```
  eval ($a213=100*(rand()-0.5) )
  eval ($a214=100*(rand()-0.5) )
  eval ($a215=100*(rand()-0.5) )
  vector do (x=x+$a213) (segid="PC")
  vector do (y=y+$a214) (segid="PC")
  vector do (z=z+$a215) (segid="PC")
  vector do (vx=50) (segid="PC")
  vector do (vy=50) (segid="PC")
  vector do (vz=50) (segid="PC")
  set disp=coor.dat end
  display cycle $a01: impulse: $a213, $a214, $a215 to x,y,z; v=50
  set disp=OUTPUT end
 eval ($a18=0)
 eval ($min2=9999.0)
 end if
 eval ($a18=$a18+1)
  if ($TEMP>200000.0) then                    !correct excessive temperature
    eval ($a45=RAND()+0.01)
    vector do (vx=$a56*$a45*rand()) (segid="PC")
    vector do (vy=$a56*$a45*rand()) (segid="PC")
    vector do (vz=$a56*$a45*rand()) (segid="PC")
   end if
   eval ($a01 = $a01 + 1)
 end loop calc
if ($ref > 0) then
  vector idend ( store9 ) ( name ca or name n or name c )        !backbone selection
  eval ($ref_file="structure_"+encode($ref)+".pdb")
  set display=rms.dat end
  display Backbone pairwise RMSD from the lowest energy strucutre ($ref_file)
  display  file           rmsd
  display ------------------------------------------------------------
  set disp=OUTPUT end
    coor swap end
    coor init end
    coor swap end
    coor disp=comp @@$ref_file
  eval ($count1=0)
  while ($count1 < $nout1) loop fill
   evaluate ($count1=$count1+1)
        evaluate ($file="structure_"+encode($count1)+".pdb")
        coor init end
        coor @@$file
        coor sele=(recall 9) fit end
        coor sele=(recall 9) rms end
        eval ($b1 = $result)
   set display=rms.dat end
   display  $file           $b1
   end loop fill
end if
set display=OUTPUT end
 set echo=true end
  stop
```

Appendix 4

Input file for rigid-body docking of Cyt *f*-Pc using PCS restraints for Xplor-NIH 2.9.9

```
parameter
        @nostoc_cytf_SL_pc.par
end
structure
        @nostoc_cytf_SL_pc.psf
end
eval ($inifile="nostoc_cytf_SL_pc.pdb")
coordinates @$inifile
@learn.pcf.par
constraints fix (segid="CYTF" or resn SL) end
eval ($a05 = 6000)                      !total nr cycles / $a53
eval ($a53 = 1000)                      !nr of steps
eval ($u2 = 30)                 !nr of cycles per docking approach
eval ($a54 = 0.01)              !timestep in ps
eval ($a61 = $a54*$a53)         !time per cycle (ps)
eval ($a80 = $a61*$a05)         !total time (ps)
eval ($a58=30.0)                !velocity factor
eval ($a56=1.5)                 !fbeta
eval ($a55=300.0)               !TBATH
eval ($a59=45.0)                !.pdb writing threshold
eval ($van=100.0)               ! vdw writing threshold
eval ($a59a=0.01*$a59)                  !vx-reset minimum
eval ($a60=1.0)                 !.pdb lag factor
eval ($a301=1.0)                !VDW repel scale factor
eval ($a93=1)                   !time-spent factor: 0<$a93<1 (fraction bound)
eval ($g1a = "N71C")            !names of the spin label 1 used in calculation
eval ($g1b = "Q104C")           !names of the spin label 2 used in calculation
eval ($g1c = "S192C")           !names of the spin label 3 used in calculation
eval ($g2a = 13)     !residue number of the 1stSL conformation of SL 1
eval ($g2b = 17)     !residue number of the 1st SL conformation of SL 2
eval ($g2c = 21)     !residue number of the 1st SL conformation of SL 3
eval ($field=600.1328E6)        !field strength in Hz
eval ($frl=1)                           !fraction of proteins with spin label
eval ($tau_c=30.0E-9)                   !tau c for the complex in sec
eval ($a9=0.01)                         !general scaling PRE (NOE) term
eval ($a10=1.0)      !scale factor for CL1 (peaks disappeared)
eval ($a11=1.0)                 !scale factor for CL2 (peaks unaffected)
eval ($a12=1.0)                 !scale factor for CL3 (peaks reduced)
eval ($low1=14.0)  !lower(d_minus) limit (restraints upper limit only)
eval ($up1=4.0)    !upper(d_plus) limit (restraints upper limit only)
eval ($low2=4.0)     !lower(d_minus) limit (restraints lower limit only)
eval ($up2=100.0) !upper(d_plus) limit (restraints lower limit only)
eval ($low3=4.0)   !lower(d_minus) limit (restraints with both limits)
eval ($up3=4.0)    !upper(d_plus) limit (restraints with both limits)
eval ($a01 = 1)                 !cycle counter
eval ($a48=$cpu*1e4)
set seed=$a48 end
eval ($a14=0)
eval ($a18=1)
eval ($nout1=0)      ! Number of strucutres output per run
eval ($min1=9999.0)
eval ($min2=9999.0)
eval ($ref=0)
set display=coor.dat end                !write parameters
display xx ----------------------PARAMETERS---------------------------
display xx startdate:        $DATE
display xx starttime:        $TIME
```

```
 display xx ini.file:            $inifile
 display xx timestep (ps):     $a54
 display xx nr of steps:       $a53              time /cycle (ps):   $a61
 display xx nr of cycles:      $a05              total time  (ps):   $a80
 display xx time-spent fact:   $a93
 display xx general scale:     $a9
 display xx TBATH:      $a55    velocity factor:   $a56   fbeta:   $a58
 display xx .pdb threshold     $a59
 display xx .pdb lag factor    $a60
 display xx Local mimim. impulse after 10 cycles with 'constant' Etot > $a59a
 display xx ------------------------------------------------------------
 display
 set display=OUTPUT end
 flag exclude elec bond angl dihe impr include vdw noe end
 vector do (fbeta=$a58) (segid="PC")
 vector do (vx=$a56) (segid="PC")
 vector do (vy=$a56) (segid="PC")
 vector do (vz=$a56) (segid="PC")
set disp=ener.dat end
display Energies for the output structures
display file                Etot          VDW         NOE
display ------------------------------------------------
set disp=OUTPUT end
@restraints.xpl
eval ($a213=100*(rand()-0.5) ) ! push away PC
eval ($a214=100*(rand()-0.5) )
eval ($a215=100*(rand()-0.5) )
vector do (x=x+$a213) (segid="PC")
vector do (y=y+$a214) (segid="PC")
vector do (z=z+$a215) (segid="PC")
!write coor OUTPUT="test1.pdb" end
parameter
     @nbfix.4sl_expl.xpl
      nbonds
            cutnb=8.5
            inhi=0.25
            ctofnb=7.5
            ctonnb=6.5
            repe1=0.6 !0.6
                  NBXMod=-2
            rexp=2
            irex=2
            rcon=$a301
            wmin=1.5
       end
 end
 constraints
             interactions (segid="CYTF")(segid="PC")
 end
 energy end
 while ($a01 LE $a05) loop calc
  display cycle $a01
  dynamics rigid
              dt=$a54
              group=(segid="PC")
              dynmode=TCOU
              tbath=$a55
              nprint=500
              nstep=$a53
              NTRFRQ=0 !new for XPLOR vs 3.8
  end
 eval ($a14=$ENER)
 if ($a14 < $min2) then
   coor copy end
```

```
  eval ($min2 = $a14)
  eval ($abc=$VDW)
  eval ($a15=$NOE)
  eval ($a16=$a01)
  eval ($a17=$a18)
end if
if ($a18 = $u2) then                !counting number of dockings
 if ($min2 < $a59) then
  if ($abc < $van) then
            coor swap end
            eval ($nout1=$nout1+1)
            eval ($pdb="structure_"+encode($nout1)+".pdb")
            write coord output=$pdb end
            if ($a14<$min1) then
             eval ($ref=$nout1)
       eval ($min1 = $a14)
      end if
      pick bond (segid="CYTF" and name FE) (segid="PC" and name CU) geom
      eval ($a50=$RESULT)
      set disp=ener.dat end
      display $nout1  $min2      $abc       $a15
      set disp=OUTPUT end
      set disp=coor.dat end
      display Cycle: $a16         Dock step: $a17      File#: $nout1          Fe-Cu: $a50             temp: $TEMP
      display  Etot:    $min2                  Evdw:  $abc             Enoe:  $a15
      display ------------------------------------------------------------
            set disp=OUTPUT end
    !violation analysis
            eval ($b3="structure_"+encode($nout1)+".viol.dat")
            set display=$b3 end
            display Violation analysis PRE data
            display res#   PRE(meas)  PRE(calc)  Dist(calc)
            display
            for $g2 in ($g2a $g2b $g2c)
            loop spinlabel
              eval ($g2d=$g2)
              eval ($g2e=$g2+1)
              eval ($g2f=$g2+2)
              eval ($g2g=$g2+3)
              if ($g2=$g2a) then
                eval ($g1h = $g1a)
                display Spin Label $g1h
                for
                        $c02 in ID (store1)
                loop C3
                  vector show elem (resi) (ID $c02)
                  eval ($c04=$RESULT)                                   !residue nr
                  vector show elem (store1) (ID $c02)
                  eval ($c08=$RESULT)
                  pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
                  eval ($ca1=$RESULT)
                  pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
                  eval ($ca2=$RESULT)
                  pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
                  eval ($ca3=$RESULT)
                  pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
                  eval ($ca4=$RESULT)
                  eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                                !r-6 average distance over 4 positions
                  eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                                !back calculated PRE
                  eval ($c10=$c03^(-1/6))
                  display $c04       $c08        $c05         $c10
                end loop C3
```

```
      display
     end if
     if ($g2=$g2b) then
      eval ($g1h = $g1b)
      display Spin Label $g1h
      for
            $c02 in ID (store2)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store2) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
     if ($g2=$g2c) then
      eval ($g1h = $g1c)
      display Spin Label $g1h
      for
            $c02 in ID (store3)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store3) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
  end loop spinlabel
      close $b3 end
      set display=OUTPUT end
 !end violoation analysis
  coor swap end
 end if
end if
eval ($a213=100*(rand()-0.5) ) !impulse to escape local minimum
```

```
  eval ($a214=100*(rand()-0.5) )
  eval ($a215=100*(rand()-0.5) )
  vector do (x=x+$a213) (segid="PC")
  vector do (y=y+$a214) (segid="PC")
  vector do (z=z+$a215) (segid="PC")
  vector do (vx=50) (segid="PC")
  vector do (vy=50) (segid="PC")
  vector do (vz=50) (segid="PC")
  set disp=coor.dat end
  display cycle $a01: impulse: $a213, $a214, $a215 to x,y,z; v=50
  set disp=OUTPUT end
 eval ($a18=0)
 eval ($min2=9999.0)
 end if
 eval ($a18=$a18+1)
  if ($TEMP>200000.0) then                    !correct excessive temperature
    eval ($a45=RAND()+0.01)
    vector do (vx=$a56*$a45*rand()) (segid="PC")
    vector do (vy=$a56*$a45*rand()) (segid="PC")
    vector do (vz=$a56*$a45*rand()) (segid="PC")
   end if
   eval ($a01 = $a01 + 1)
 end loop calc
if ($ref > 0) then
  vector idend ( store9 ) ( name ca or name n or name c )        !backbone selection
  eval ($ref_file="structure_"+encode($ref)+".pdb")
  set display=rms.dat end
  display Backbone pairwise RMSD from the lowest energy strucutre ($ref_file)
  display  file           rmsd
  display ------------------------------------------------------------
  set disp=OUTPUT end
    coor swap end
    coor init end
    coor swap end
    coor disp=comp @@$ref_file
  eval ($count1=0)
  while ($count1 < $nout1) loop fill
   evaluate ($count1=$count1+1)
        evaluate ($file="structure_"+encode($count1)+".pdb")
        coor init end
        coor @@$file
        coor sele=(recall 9) fit end
        coor sele=(recall 9) rms end
        eval ($b1 = $result)
   set display=rms.dat end
   display  $file           $b1
   end loop fill
end if
set display=OUTPUT end
 set echo=true end
  stop
```

Appendix 5

Input file to backcalculate PCS from Cyt *f* Fe to Pc amide protons for Xplor-NIH 2.9.9

```
parameter
        @nostoc_cytf_SL_pc.par
end
structure
        @nostoc_cytf_SL_pc.psf
end
eval ($inifile="nostoc_cytf_SL_pc.pdb")
coordinates @$inifile
@learn.pcf.par
constraints fix (segid="CYTF" or resn SL) end
eval ($a05 = 6000)                      !total nr cycles / $a53
eval ($a53 = 1000)                      !nr of steps
eval ($u2 = 30)                 !nr of cycles per docking approach
eval ($a54 = 0.01)              !timestep in ps
eval ($a61 = $a54*$a53)         !time per cycle (ps)
eval ($a80 = $a61*$a05)         !total time (ps)
eval ($a58=30.0)                !velocity factor
eval ($a56=1.5)                 !fbeta
eval ($a55=300.0)               !TBATH
eval ($a59=45.0)                !.pdb writing threshold
eval ($van=100.0)               ! vdw writing threshold
eval ($a59a=0.01*$a59)                  !vx-reset minimum
eval ($a60=1.0)                 !.pdb lag factor
eval ($a301=1.0)                !VDW repel scale factor
eval ($a93=1)                   !time-spent factor: 0<$a93<1 (fraction bound)
eval ($g1a = "N71C")            !names of the spin label 1 used in calculation
eval ($g1b = "Q104C")           !names of the spin label 2 used in calculation
eval ($g1c = "S192C")           !names of the spin label 3 used in calculation
eval ($g2a = 13)    !residue number of the 1stSL conformation of SL 1
eval ($g2b = 17)    !residue number of the 1st SL conformation of SL 2
eval ($g2c = 21)    !residue number of the 1st SL conformation of SL 3
eval ($field=600.1328E6)        !field strength in Hz
eval ($frl=1)                           !fraction of proteins with spin label
eval ($tau_c=30.0E-9)                   !tau c for the complex in sec
eval ($a9=0.01)                         !general scaling PRE (NOE) term
eval ($a10=1.0)     !scale factor for CL1 (peaks disappeared)
eval ($a11=1.0)                 !scale factor for CL2 (peaks unaffected)
eval ($a12=1.0)                 !scale factor for CL3 (peaks reduced)
eval ($low1=14.0)  !lower(d_minus) limit (restraints upper limit only)
eval ($up1=4.0)    !upper(d_plus) limit (restraints upper limit only)
eval ($low2=4.0)    !lower(d_minus) limit (restraints lower limit only)
eval ($up2=100.0)   !upper(d_plus) limit (restraints lower limit only)
eval ($low3=4.0)   !lower(d_minus) limit (restraints both limits)
eval ($up3=4.0)    !upper(d_plus) limit (restraints both limits)
eval ($a01 = 1)                 !cycle counter
eval ($a48=$cpu*1e4)
set seed=$a48 end
eval ($a14=0)
eval ($a18=1)
eval ($nout1=0)     ! Number of strucutres output per run
eval ($min1=9999.0)
eval ($min2=9999.0)
eval ($ref=0)
set display=coor.dat end                !write parameters
display xx ----------PARAMETERS----------
display xx startdate:       $DATE
```

```
 display xx starttime:        $TIME
 display xx ini.file:          $inifile
 display xx timestep (ps):    $a54
 display xx nr of steps:      $a53              time /cycle (ps):   $a61
 display xx nr of cycles:     $a05              total time  (ps):   $a80
 display xx time-spent fact:  $a93
 display xx general scale:    $a9
 display xx TBATH:    $a55    velocity factor:   $a56   fbeta:   $a58
 display xx .pdb threshold    $a59
 display xx .pdb lag factor   $a60
 display xx Local mimim. impulse after 10 cycles with 'constant' Etot > $a59a
 display xx ------------------------------------------------------------
 display
 set display=OUTPUT end
 flag exclude elec bond angl dihe impr include vdw noe end
 vector do (fbeta=$a58) (segid="PC")
 vector do (vx=$a56) (segid="PC")
 vector do (vy=$a56) (segid="PC")
 vector do (vz=$a56) (segid="PC")
set disp=ener.dat end
display Energies for the output structures
display file              Etot          VDW         NOE
display ----------------------------------------------------
set disp=OUTPUT end
@restraints.xpl
eval ($a213=100*(rand()-0.5) ) ! push away PC
eval ($a214=100*(rand()-0.5) )
eval ($a215=100*(rand()-0.5) )
vector do (x=x+$a213) (segid="PC")
vector do (y=y+$a214) (segid="PC")
vector do (z=z+$a215) (segid="PC")
!write coor OUTPUT="test1.pdb" end
parameter
     @nbfix.4sl_expl.xpl
      nbonds
            cutnb=8.5
            inhi=0.25
            ctofnb=7.5
            ctonnb=6.5
            repe1=0.6 !0.6
                        NBXMod=-2
            rexp=2
            irex=2
            rcon=$a301
            wmin=1.5
      end
 end
 constraints
             interactions (segid="CYTF")(segid="PC")
 end
 energy end
 while ($a01 LE $a05) loop calc
  display cycle $a01
  dynamics rigid
             dt=$a54
             group=(segid="PC")
             dynmode=TCOU
             tbath=$a55
             nprint=500
             nstep=$a53
             NTRFRQ=0 !new for XPLOR vs 3.8
  end
 eval ($a14=$ENER)
 if ($a14 < $min2) then
```

```
  coor copy end
  eval ($min2 = $a14)
  eval ($abc=$VDW)
  eval ($a15=$NOE)
  eval ($a16=$a01)
  eval ($a17=$a18)
 end if
 if ($a18 = $u2) then                !counting number of dockings
  if ($min2 < $a59) then
   if ($abc < $van) then
              coor swap end
              eval ($nout1=$nout1+1)
              eval ($pdb="structure_"+encode($nout1)+".pdb")
              write coord output=$pdb end
              if ($a14<$min1) then
               eval ($ref=$nout1)
        eval ($min1 = $a14)
       end if
       pick bond (segid="CYTF" and name FE) (segid="PC" and name CU) geom
       eval ($a50=$RESULT)
       set disp=ener.dat end
       display $nout1  $min2       $abc        $a15
       set disp=OUTPUT end
       set disp=coor.dat end
       display Cycle: $a16          Dock step: $a17       File#: $nout1            Fe-Cu: $a50              temp: $TEMP
       display  Etot:    $min2                      Evdw:  $abc              Enoe:  $a15
       display ------------------------------------------------------------
              set disp=OUTPUT end
     !violation analysis
              eval ($b3="structure_"+encode($nout1)+".viol.dat")
              set display=$b3 end
              display Violation analysis PRE data
              display res#   PRE(meas)  PRE(calc)  Dist(calc)
              display
              for $g2 in ($g2a $g2b $g2c)
              loop spinlabel
                eval ($g2d=$g2)
                eval ($g2e=$g2+1)
                eval ($g2f=$g2+2)
                eval ($g2g=$g2+3)
                if ($g2=$g2a) then
                  eval ($g1h = $g1a)
                  display Spin Label $g1h
                  for
                            $c02 in ID (store1)
                  loop C3
                    vector show elem (resi) (ID $c02)
                    eval ($c04=$RESULT)                                  !residue nr
                    vector show elem (store1) (ID $c02)
                    eval ($c08=$RESULT)
                    pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
                    eval ($ca1=$RESULT)
                    pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
                    eval ($ca2=$RESULT)
                    pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
                    eval ($ca3=$RESULT)
                    pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
                    eval ($ca4=$RESULT)
                    eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                                  !r-6 average distance over 4 positions
                    eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                                  !back calculated PRE
                    eval ($c10=$c03^(-1/6))
                    display $c04       $c08        $c05         $c10
```

```
      end loop C3
      display
     end if
     if ($g2=$g2b) then
      eval ($g1h = $g1b)
      display Spin Label $g1h
      for
            $c02 in ID (store2)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store2) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
     if ($g2=$g2c) then
      eval ($g1h = $g1c)
      display Spin Label $g1h
      for
            $c02 in ID (store3)
      loop C3
       vector show elem (resi) (ID $c02)
       eval ($c04=$RESULT)                          !residue nr
       vector show elem (store3) (ID $c02)
       eval ($c08=$RESULT)
       pick bond (resn SL and resi $g2d and name O) (ID $c02) geom
       eval ($ca1=$RESULT)
       pick bond (resn SL and resi $g2e and name O) (ID $c02) geom
       eval ($ca2=$RESULT)
       pick bond (resn SL and resi $g2f and name O) (ID $c02) geom
       eval ($ca3=$RESULT)
       pick bond (resn SL and resi $g2g and name O) (ID $c02) geom
       eval ($ca4=$RESULT)
       eval ($c03=($ca1^(-6)+$ca2^(-6)+$ca3^(-6)+$ca4^(-6))/4)
                    !r-6 average distance over 4 positions
       eval ($c05=(1.23E16*(4*$tau_c+(3*$tau_c/(1+($field*2*3.14*$tau_c)^2)))*$frl*$a93*$c03) )
                    !back calculated PRE
       eval ($c10=$c03^(-1/6))
       display $c04    $c08    $c05    $c10
      end loop C3
      display
     end if
  end loop spinlabel
      close $b3 end
      set display=OUTPUT end
 !end violoation analysis
  coor swap end
 end if
end if
```

```
                                                !impulse to escape local minimum
  eval ($a213=100*(rand()-0.5) )
  eval ($a214=100*(rand()-0.5) )
  eval ($a215=100*(rand()-0.5) )
  vector do (x=x+$a213) (segid="PC")
  vector do (y=y+$a214) (segid="PC")
  vector do (z=z+$a215) (segid="PC")
  vector do (vx=50) (segid="PC")
  vector do (vy=50) (segid="PC")
  vector do (vz=50) (segid="PC")
  set disp=coor.dat end
  display cycle $a01: impulse: $a213, $a214, $a215 to x,y,z; v=50
  set disp=OUTPUT end
 eval ($a18=0)
 eval ($min2=9999.0)
 end if
 eval ($a18=$a18+1)

  if ($TEMP>200000.0) then                      !correct excessive temperature
    eval ($a45=RAND()+0.01)
    vector do (vx=$a56*$a45*rand()) (segid="PC")
    vector do (vy=$a56*$a45*rand()) (segid="PC")
    vector do (vz=$a56*$a45*rand()) (segid="PC")
   end if
   eval ($a01 = $a01 + 1)
 end loop calc
if ($ref > 0) then
  vector idend ( store9 ) ( name ca or name n or name c )         !backbone selection
  eval ($ref_file="structure_"+encode($ref)+".pdb")
  set display=rms.dat end
  display Backbone pairwise RMSD from the lowest energy strucutre ($ref_file)
  display  file            rmsd
  display ----------------------------------------------------------------
  set disp=OUTPUT end
    coor swap end
    coor init end
    coor swap end
    coor disp=comp @@$ref_file
  eval ($count1=0)
  while ($count1 < $nout1) loop fill
   evaluate ($count1=$count1+1)
        evaluate ($file="structure_"+encode($count1)+".pdb")
        coor init end
        coor @@$file
        coor sele=(recall 9) fit end
        coor sele=(recall 9) rms end
        eval ($b1 = $result)
   set display=rms.dat end
   display  $file           $b1
   end loop fill
end if
set display=OUTPUT end
 set echo=true end
  stop
```

List of publications

Scanu S., Förster J.M., Timmer M., Ullmann G.M. and Ubbink M. (2013) Loss of electrostatic interactions causes increase of dynamics within the plastocyanin-cytochrome *f* complex. *Manuscript under revision.*

Scanu S., Förster J.M., Ullmann G.M. and Ubbink M. (2013) Role of hydrophobic interactions in the encounter complex formation of plastocyanin and cytochrome *f* complex revealed by paramagnetic NMR spectroscopy. *J. Am. Chem. Soc.* **135**, 7681-7692.

Scanu S., Förster J., Finiguerra M.G., Shabestari M.H., Huber M. and Ubbink M. (2012) The complex of cytochrome *f* and plastocyanin from *Nostoc sp.* PCC 7119 is highly dynamic. *Chembiochem*. **13**, 1312-1318

Bashir Q., **Scanu S.**, and Ubbink M. (2011) Dynamics in electron transfer protein complexes. *FEBS J*. **278**, 1391-1400

Xu X., **Scanu S.**, Chung J.S., Hirasawa M., Knaff D.B. and Ubbink M. (2010) Structural and functional characterization of the Ga-substituted ferrodoxin from *Synecocystis* sp. PCC6803, a mimic of the native protein. *Biochemistry*, **49**, 7790-7797

Barbiroli, A., Beringhelli, T., Bonomi, F., Donghi, D., Ferranti, P., Galliano, M., Iametti, S., Maggioni, D., Rasmussen, P., **Scanu, S.** and Vilardo, M. C. (2010) Bovine β-lactoglobulin acts as an acid-resistan drug carrier by exploiting its diverse binding regions. *Biological Chemistry*. **391**, 21-32

Curriculum vitae

Sandra Scanu was born on August 27th 1981 in Pattada, a small village located in the heart of Sardinia. She spent a light-hearted childhood in contact with wild and beautiful nature, surrounded by the love of family and friends. Since she was a child, she has showed her curiosity for the world by asking unexpected and, sometimes, surprising questions about life and people. After obtaining her diploma in the scientific high school A. Segni in 2000, she decided to move to Milan to pursue her dream of becoming a scientist. Here, she enrolled in Medical Biotechnology and she obtained her Bachelor degree in 2004, after attending a three months internship in the lab of Prof. G.P. Comi at the Policlinico of Milan. After one year of Chemistry courses, she started her studies in Pharmaceutical Biotechnology, earning a master's degree in 2008. From 2006 to 2008 she worked in the lab of Prof. T. Beringhelli, first as a master's student then as a post-graduate fellow. During these years she has remained fascinated by the endless applications of NMR spectroscopy. This experience led her to Leiden in 2008, where she became a PhD student in the ProtChem group under the supervision of Prof. M. Ubbink. Here, she studied the principles of protein interactions using paramagnetic NMR spectroscopy. Once she overcame the initial troubles with English, she managed to settle down perfectly in the amazing city that accommodated her for four years. Soon she met friends from all over the world, with whom she organized many gatherings and dinner parties. Beside her passion for science, Sandra is in fact successfully dedicated to another art: the good cuisine.
She is now about to begin another journey to Munich, where she will start a post-doctorate position in the group of Biomolecular NMR spectroscopy under the supervision of Dr. T. Madl.

By Giannina Scanu