NMR studies of protein-small molecule and protein-peptide interactions
Guan, J.

Cover Page

Universiteit Leiden

Author: Guan, Jia-Ying
Title: NMR studies of protein-small molecule and protein-peptide interactions
Issue Date: 2013-12-02

NMR studies of protein-small molecule and protein-peptide interactions

Proefschrift

te 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 maandag 2 december 2013

klokke 13:45 uur

door

Jia-Ying Guan

Geboren te Yi-Lan, Taiwan in 1980

Promotiecommissie

Promotor: Prof. Dr. M. Ubbink

Co-promotor: Dr. G. Siegal

Overige leden: Prof. Dr. J. Brouwer (Universiteit Leiden)

Prof. Dr. H. van Vlijmen (Universiteit Leiden)

Prof. Dr. J. van Duynhoven (Universiteit Wageningen)

Prof. Dr. C. Luchinat (University of Florence)

Dr. M. van der Stelt (Universiteit Leiden)

Front cover: Turtle island (龜山島), landmark of Yi-Lan, Taiwan. Original photo was kindly provided by Min-Jen Lin (林明仁), a photographer from Yi-Lan, Taiwan.

Rear cover: Oudenhofmolen, landmark of Oegstgeest, the Netherlands.

To my beloved families

who always support me in my decisions

Table of Contents

List of Abbreviations

|	
CLaNP	Caged Lanthanide NMR Probe
CPMG	Carr-Purcell-Meiboom-Gill
CSP	chemical shift perturbations
Cyt f	Cytochrome f
DNase	deoxyribonuclease
DTT	dithiotheitol
EPR	electron paramagnetic resonance
FBDD	fragment-based drug discovery
FKBP12	FK506-binding protein 12
HSQC	heteronuclear single-quantum correlation spectroscopy
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_D	dissociation constant
LBP	lanthanide-binding peptide
LBT	lanthanide-binding tag
MC	Monte Carlo
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser enhancement spectroscopy
Pc	plastocyanin
PCS	pseudocontact shift
PMSF	phenylmethylsulfonyl fluoride
PRE	paramagnetic relaxation enhancement
RDC	residual dipolar coupling
SAR	structure-activity relationship
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TFA	trifluoroacetid acid
TINS	target-immobilized NMR screening
TOAC	2,2,6,6-tetramethylpiperidine-1-oxyl-4.amino-4-carboxylic acid
TOCSY	total correlation spectroscopy

1 General introduction

1.1 Protein-ligand interactions

Protein-ligand interactions are essential for all biological processes. These interactions comprise biological recognition at the molecular level. When the ligand binds to the protein, it can regulate binding of other molecules, either by directly disrupting the interaction, competition of binding, or indirectly causing conformational changes. Therefore, understanding the detailed interactions between the binding partners is important to gain insight into the corresponding biological processes. This thesis focuses on two proteins: 12 kDa FK-506 binding protein (FKBP12) and plastocyanin (Pc). FKBP12 is used as a model protein for the development of a paramagnetic NMR-based methodology for studying general protein-small molecule interactions. The transient complexes formed by three different Pcs and small charged peptides, are created as models to elucidate the dynamic nature of encounter complexes and to provide fundamental understanding of transient protein-protein interactions.

FK-506 binding protein (FKBP)

FKBP is a protein family with *cis-trans* peptidyl-prolyl isomerase activity and belongs to the immunophilin family. FKBPs have been identified in many eukaryotes and function as protein folding chaperones for proteins containing proline residues. In this family, FKBP12 is a ubiquitous cytosolic protein implicated in several physiological processes. It is known in humans for binding the immunosuppressant molecules FK506 (tacrolimus)[1,2] and rapamycin (sirolimus),[3] which are used in treating patients after organ transplant and patients suffering from autoimmune disorders. The FKBP12-FK506 complex inhibits a serine/threonine phosphatase, calcineurin,[4] thus blocking the T-lymphocyte signal transduction pathway.[5,6]
The structures of FKBP12 in the free form, with ligands or other proteins, have been solved by NMR spectroscopy[7–14] and X-ray crystallography.[15–29] The overall structure is made up of five β-strands, an α-helix, and three loops.[15] The structure indicates a unique ligand binding

pocket, which can be subdivided into two sites.[30] Site 1 is the binding location for the pipecolinic acid ligands, whereas site 2 lies about 7 Å away. Selected ligands with a range of affinities for sites 1 or 2 have been reported, some of which have been linked to generate high-affinity ligands (Figure 1.1). This study pioneered the field of fragment-based drug discovery.
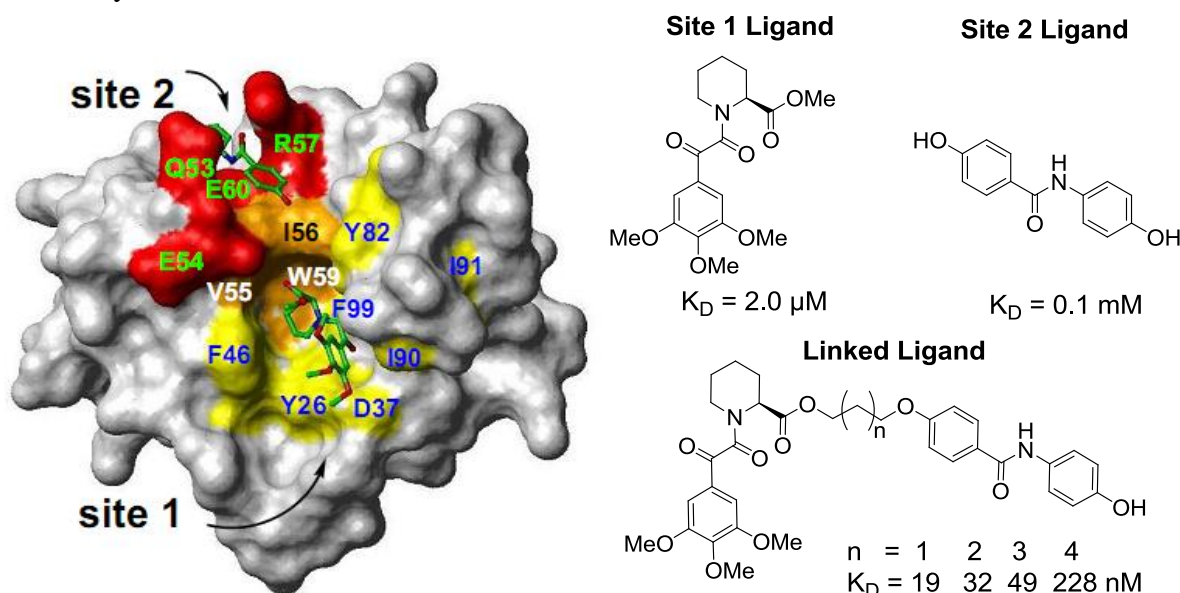
Figure 1.1: (Left) Surface representation of the two binding sites of FKBP12 with site-specific ligands. Residues in yellow and red are in site 1 and site 2, respectively. Residues in orange are shared between the two sites. For the ligands, atoms in red, blue and green are oxygens, nitrogens and carbons, respectively. (Right) Site-specific FKBP ligands are linked to produce high affinity ligands.[30]

Plastocyanin

Plastocyanin (Pc) is a blue copper protein which is involved in the electron transport process in photosynthesis. It transports electrons between cytochrome f of the b_6f complex and $P700^+$ of photosystem I (PSI) in plants, green algae and cyanobacteria.[31–34] Pcs from different species have been extensively studied and their structures are available from various plants and bacteria.[35–49] The overall structures of Pc are highly conserved from cyanobacteria to higher plants,[42] ranging from 97 to 105 amino acid residues, but large differences are present in terms of surface charge distributions. The typical structure of Pc consists of a β-sandwich consisting of seven β-strands and an irregular strand which contains a small helix in some Pcs. In poplar Pc, for example, the copper ion is chelated strongly by 3 residues, namely the S^γ of C84, the N^δ atoms of H37 and H87 and weakly with the S^δ of M92 (Figure 1.2).[50]
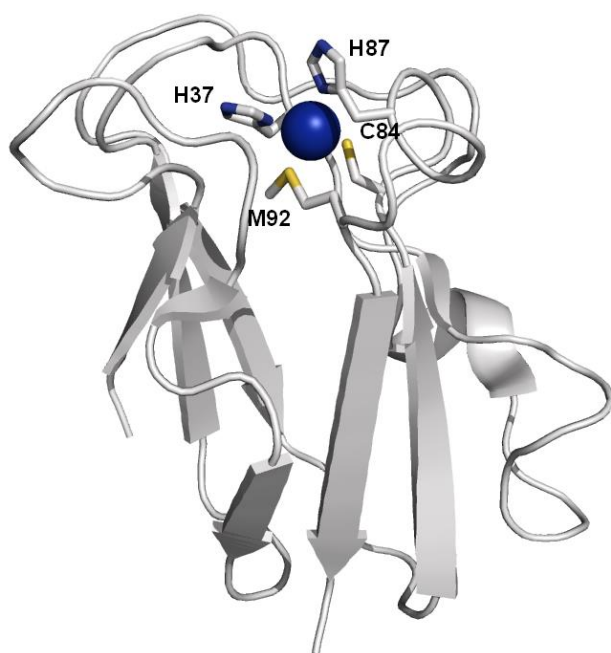
Figure 1.2: Structure of poplar Pc (PDB entry 1TKW).[47] Pc is shown as a grey ribbon, with the Cu ion as a blue sphere. The four residues (H37, C84, H87 and M92) that serve as the copper ligands are shown in sticks.

The function of Pc relies on two sites: a hydrophobic patch surrounding the solvent-accessible histidine copper ligand, and an electrostatically charged, remote surface area, whose nature varies from one organism to another. This charged area is mainly acidic in plants and green algae, whereas it ranges from acidic to basic in cyanobacteria. In higher plants, the hydrophobic patch surrounding the exposed His87 ligand and the distant acidic patch are potential binding sites for redox partners of Pc.[41,51–54] In the fern *Dryopteris crassirhizoma* Pc, on the contrary, the acidic region surrounds the hydrophobic patch.[43] Although the acidic patches are not conserved in bacteria, the hydrophobic patch is always present. The hydrophobic patches and the acidic patches play a key role in the dynamics of the complex.[51]

1.2   Fragment-based drug discovery

Fragment-based drug discovery (FBDD) is a powerful method for discovering high-affinity ligands for target proteins. The term "fragment" is used for small organic molecules with a molecular weight of <300 Da.[55] Fragment-based screening has several advantages over conventional high-throughput screening (HTS). First, due to the small size of fragments, the number of compounds that need to be screened is dramatically reduced to cover a greater chemical diversity space. Second, although fragment hits are weakly binding, they make high-quality interactions with the target to bind with sufficient affinity for detection, which has been interpreted as high ligand efficiency.[56]

 In FBDD, detection of weak binding is of particularly interest, which leads to the development of various biophysical tools. Overview of various techniques applied in FBDD can be found elsewhere.[57–59] The following content describes three of the most commonly used techniques to characterize ligand binding. An overview of existing protein-ligand NMR approaches will be presented in Section 1.3.

X-ray crystallography

X-ray crystallography ins an information-rich technique that provides both hit validation and structural information in one step. It is often accomplished by analyzing protein crystals soaked in a cocktail of fragments.[60] This technique is a versatile tool to visualize the complex structures at the atomic level, although it is not possible to measure an affinity. However, many cases have shown failures in obtaining good quality crystals due to problems of ligand soaking, multiple binding modes, weak binding, or disruption of the crystal lattice caused by protein motions. It is also a huge investment of resource to obtain the simple yes/no answer in primary screening.

Surface plasmon resonance

Surface plasmon resonance (SPR) is a technique based on optical sensors. It provides binding measurements to characterize the kinetic properties of the biomolecular complex. In SPR, the target protein is immobilized on a sensor chip with a thin layer of gold on the surface, which is then exposed to the analyte in the flow channel.[61] Binding of the ligands results in changes of the refractive index on the surface of the sensor. This change is proportional to the number of bound molecules. The rate constants of ligand association (k_{on}) and ligand dissociation (k_{off}) are estimated by regression of the association and dissociation gradients at different ligand concentrations. The ratio of k_{off}/k_{on} is the binding constant (K_D). Therefore, SPR can quantitatively analyze the binding events.[62]

Isothermal titration calorimetry

Recently, thermodynamics involved in ligand binding has gained much interest in optimizing the structures of drug candidates.[63,64] Isothermal titration calorimetry (ITC) is one of the few techniques which measure both the binding affinity and the thermodynamic terms that contribute to the affinity: the binding enthalpy (ΔH) and the binding entropy (ΔS). It has no requirement for chemical modification, labeling, immobilization or limit on the size of interacting species. The binding curve allows extraction of the thermodynamic parameters, the stoichiometry and K_D. This technique has been used to provide a comparison of energy terms in the development of complete series of drugs.[63] To achieve high affinity, both ΔH and ΔS are required to contribute favorably to binding. Therefore, the thermodynamic data can provide valuable information to complement existing tools to facilitate lead discovery and optimization.[64,65]

1.3 NMR techniques for protein-ligand interactions

NMR has been recognized as a powerful tool to characterize macromolecular structures and dynamics. It is also utilized extensively in screening ligand binding to protein targets in the early stage of fragment-based drug discovery.[66–70] One of its key advantages is that it can detect and quantify weak interactions (K_D μM-mM) without prior knowledge of protein structures. In addition, NMR can be applied to obtain structural information for both the protein target and the ligand with atomic resolution for subsequent optimization of weak binding hits into high affinity drug candidates.

Here, a summary of NMR techniques used for protein-ligand interactions is presented. They are subdivided into ligand- and protein-observed techniques. In the protein-observed techniques, the influences of ligand binding on the spectra of the target are detected, whereas in the ligand-observed techniques, changes induced in the ligand's NMR resonances upon binding are observed. Recently, a variety of techniques using paramagnetic NMR and combining NMR and computational approaches are gaining interests. These techniques are also described here.

Protein-observed NMR

Binding of ligands to targets leads to changes in the chemical shifts, which are most conveniently monitored in mixtures of ligands with a ^{15}N- or ^{13}C-labeled protein. Through a series of target 2D-HSQC spectra with different ligand concentrations, the dissociation constant K_D can be derived. If the assignments of the target are available, the location of the binding site may be easily derived. This strategy is called chemical shift perturbation (CSP), also known as SAR (Structure-Activity Relationship) by NMR,[30] chemical shift mapping or complexation-induced change in chemical shift (CIS), and has been successful applied in numerous cases for screening and optimizing lead compounds. However, HSQC-based methods are presently limited to macromolecular complexes with a molecular weight below 30 or 100 kDa (in combination with TROSY).[71,72] Another limitation is the requirement for large amounts of isotopically labelled protein, which can only be obtained if an effective expression system in an isotope-labelled medium is available. Despite this, side-chain specific isotope labelling has significantly increased the scope of protein systems that are amenable for SAR by NMR,[73,74] and the cost-effective ^{13}C labelling has made screening of 100,000 compounds practical.[75] To obtain information of ligand binding at atomic resolution, intermolecular NOEs can be observed using isotope filtered/edited NOESY experiments.

This works best for tight-binding ligands. Transferred NOE experiments can be used for weak binding ligands.

Ligand-observed NMR

Another category of techniques is based on the changes in the ligand NMR signals in the presence of the target. These methods are often used for proteins with Mw 15–100 kDa. Given that the ligand is in fast exchange on the NMR time scale, the observed signals are the weighted average of the free and bound ligand. Most techniques used for primary screening detect a response in the form of a reduction (disappearance) or increase (appearance) of signal amplitude due to changes in relaxation rates or magnetization transfer. Competition/displacement experiments can also provide information of the binding site and the relative affinity if the location and affinity of the first ligand is known. Paramagnetic labelling on the ligand or the protein can also be used to obtain information of the ligand binding by monitoring changes in the relaxation or chemical shifts of nuclei located at different regions of the molecule. An overview of the techniques is shown in Table 1.1.

Saturation transfer difference (STD) was initially designed to identify small-molecule ligands that bind to a receptor protein from a mixture of compounds.[76] The spectrum of the receptor is saturated by selective irradiation that does not influence the ligand resonances. Spin diffusion in the receptor leads to saturation of its spectrum and causes saturation transfer by NOE to the bound ligand. If the ligand is in fast exchange, signal reduction of the free ligand can be detected. This method has also been applied to identify the ligand binding epitope[77] and the structure of a bound ligand.[78,79]

With the combination of STD and highly deuterated proteins with selectively protonated amino acids (Ile, Leu, Val, and Met), one can also identify ligand binding sites. This is used in SOS-NMR (*s*tructural information using *O*verhauser effects and *s*elective labelling).[80] It requires an unlabeled protein as the positive control and a perdeuterated protein as the negative control.

WaterLOGSY (water-ligand observed via gradient spectroscopy)[81–83] is based on the intermolecular magnetization transfer and spin diffusion of the excited bulk water. Small molecules that interact with the 'bound' water on the protein (binders) behave differently from those that interact with the bulk water (non-binders), resulting in different signs in NMR signals. WaterLOGSY has been widely applied in primary screening.

Target immobilized NMR screening (TINS) uses a solid support to immobilize the target and therefore slows down the tumbling rate of the immobilized target, causing much faster

relaxation relative to the free molecules in the solution phase. TINS uses difference in the spectrum of the fragments in the presence of the target to detect binding, and of a reference protein to cancel out non-specific binding. It was first demonstrated with FKBP12,[84] and has been successfully applied on membrane proteins DsbB[85] and adenosine A_{2A} receptor.[86,87]

INPHARMA (*i*nterligand *N*OEs for *phar*macophore *ma*pping)[88] detects protein-mediated ligand-ligand interactions in competition binding. The restraints come from protein-mediated inter-ligand nuclear Overhauser effects (NOEs) observed between two competitive ligands. If the structure of the ligand–protein complex is known for one ligand, the complex structure for the other ligand can be estimated based on the intermolecular NOEs.

Paramagnetic NMR

Paramagnetic effects originate from unpaired electrons, which can affect the magnetic properties of nuclei in the vicinity. The theory of paramagnetic effects is explained in Section 1.4. These long-range effects are strongly distance and, in some cases, orientation dependent, and are subsequently observable in NMR spectra as pseudocontact shifts (PCS), paramagnetic relaxation enhancement (PRE) and/or residual dipolar coupling (RDC).[89–91] Therefore, paramagnetic NMR can provide valuable information for elucidation of protein structure, conformational dynamics, protein-protein interactions and protein-ligand interactions. Here the existing paramagnetic NMR-based approaches for protein-ligand complexes are described.

SLAPSTIC (*s*pin *l*abels *a*ttached to *p*rotein *s*ide chains as a *t*ool to identify *i*nteracting *c*ompounds) uses spin-labels, such as 2,2,6,6-tetramethyl-piperidine-1-oxyl (TEMPO), attached on the lysine residues of the target protein to detect ligand binding from a mixture of compounds.[92] The bound ligand shows signal reduction in $T_{1\rho}$ relaxation experiments. This approach can be used for primary screening. A similar approach has been applied in screening a second ligand that binds close to an already identified primary ligand attached with a TEMPO spin label.[93] Such approaches have been demonstrated to find inhibitors for Bcl-x_L[93] and Bcr-Abl kinase.[94] The use of paramagnetic ions instead of spin labels for ligand screening has also been demonstrated by substitution of diamagnetic Zn^{2+} with paramagnetic Co^{2+} in the matrix metalloproteinase MMP-12.[95]

To further determine the structure of the protein-ligand complex using paramagnetic effects, several studies have appeared in recent years. A combination of PCS and RDC with respect to a single fusion C-terminus lanthanide-binding peptide tag (LBT) was applied to assist structure determination of a carbohydrate-protein complex.[96] Later, the use of PCS was

applied to determine the pose of a ligand bound to a metalloprotein.[97] This requires an enzyme with a native metal chelating site near the binding pocket, and the use of very strong lanthanides. PRE assisted ligand docking with a spin-label peptide bound specifically to a protein was also reported.[98] Recently, a two-point anchored N-terminus lanthanide binding peptide (LBP) was applied for both ligand screening and determination of protein-peptide complexes.[99] In this approach, ligand screening was achieved by observing paramagnetic line broadening for both the target protein and its bound ligand in 1D-^{1}H NMR, and the ligand binding modes were derived by paramagnetic pseudocontact shifts (PCS).

Combination of NMR and computational approaches

The advantages of computation over experimental approaches are the fast speed and the possibility of studying challenging complexes that are difficult to study experimentally. However, computational approaches do not always provide accurate prediction. By integrating the experimental information, computational studies can provide more valuable biological insight. Various approaches combining NMR data and computation have been developed to extract the information of ligand binding.[100–102]

Typically, if no significant conformational changes occurs on the protein upon ligand binding, the ring current effect generated by aromatic rings in ligands is the most important factor of the observed CSPs on the protein.[103] Detailed calculations of ligand-induced CSPs were demonstrated useful to determine ligand binding modes.[104–106] However, the quantitative CSP methods are limited to aromatic ligands, due to the requirement of ring current shifts produced by the ligand.[104] Qualitative CSPs can also be used to both guide and filter the calculations, reducing the grid space and minimized energy in the process.[107]

Protein-ligand NOE matching approach[108] utilizes data from 3D ^{13}C-edited, ^{13}C/ ^{15}N-filtered HSQC-NOESY spectra for evaluating ligand binding poses. Only the ^{1}H NMR assignments of the bound ligand are essential. No protein assignments are required but can be used if the information is available. Predicted filtered NMR spectra are used and, in combination of ligand pose evaluation, refined to match with the experimental spectra. This method required a single protein sample which has to be ^{13}C- and ^{15}N-labelled.

Table 1.1: Overview of NMR techniques for studying protein-ligand interactions

Type	Technique	Principles and applications	Limitation
Magnetization transfer	STD	Primary screening and epitope mapping.	Large protein (>10 kDa) preferred.
	waterLOGSY	Primary screening; good for very hydrophilic targets and ligands.	Large protein (>10 kDa) preferred. Labile protons can give false positives.
	SOS-NMR	STD-based; observe differences between protonated and deuterated proteins	Requires selective amino acid labeling.
	INPHARMA	NOE-based, detect inter-ligand NOEs; active site and bound ligand conformation.	Observation depends on kinetics and ratio of ligand affinities.
Relaxation/ Paramagnetic NMR	TINS	Bound ligand has enhanced R_2 due to immobilization on the resin; primary screening; applicable to membrane proteins.	Not applicable to tight-binding ligands.
	SLAPSTIC	Attach TEMPO on lysine residues; detect line-broadening of bound ligand by $T_{1\rho}$; primary screening.	Results can be ambiguous if the lysine position is unknown.
	Lanthanide binding peptides (LBP)	LBP co-expressed with the protein; screening and binding pose information.	^{15}N-labeled proteins required to determine tensor parameters.
NMR driven Computation	CSP guided docking	CSP-based; quantitatively calculate ring current induced shifts from the ligand or qualitatively guide docking; ligand binding mode.	Quantitative methods limited to aromatic ligands. ^{13}C/^{15}N-labeled proteins required.
	Protein-ligand NOE matching	NOE-based; only need ligand assignments; protein assignments not essential; ligand binding mode.	^{13}C/^{15}N-labeled proteins required.

1.4 Theory of paramagnetic effects

Due to its isotropic and anisotropic effects, paramagnetic NMR can be a versatile tool to extract position-specific distance and angular information with respect to the paramagnetic center. Subsequently, the information can be used to determine and refine the structures of protein-ligand and protein-protein complexes, facilitate resonance assignments, and to study the sparsely-populated states and the dynamics.[89,90,109–112] The three most commonly used paramagnetic effects are paramagnetic relaxation enhancement (PRE), pseudocontact shift (PCS) and residual dipolar coupling (RDC). The principles of paramagnetic effects applied in this thesis, PRE and PCS, are explained below. Detailed theories of paramagnetic effects, including RDC, can be found in literature reviews and books.[89,113–119]

Paramagnetic relaxation enhancement (PRE)

Paramagnetic centers invariably cause line broadening in the NMR spectrum owing to paramagnetic enhancements of the transverse relaxation rate, R_2. This effect is called paramagnetic relaxation enhancement (PRE). It shows a distance dependency of r^{-6}. The relation between the nucleus-to-electron distance (r_{IM}) and relaxation rate enhancement (R_2^{para}) is shown in its simplified form in Equation 1.1:

$$R_2^{para} = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r_{IM}^6}\left(4\tau_c + \frac{3\tau_c}{1+\omega_I^2\tau_c^2}\right) \tag{1.1}$$

Where μ_0 is the permeability of free space, g_e is the electronic g-factor, μ_B is the Bohr magneton, S is the total electron spin quantum number, γ_I is the proton gyromagnetic ratio, $\omega_I/2\pi$ is the Larmor frequency of the proton, and τ_c is the correlation time.

R_2^{para} can be retrieved by comparing the peak intensities in the paramagnetic and diamagnetic forms (Equation 1.2):

$$\frac{I_{para}}{I_{dia}} = \frac{R_2^{dia}\exp\left(-R_2^{para}t\right)}{R_2^{dia} + R_2^{para}} \tag{1.2}$$

where I_{para} and I_{dia} are the peak heights under paramagnetic and diamagnetic conditions, respectively, and t is the INEPT time in a HSQC experiment, the time during which transverse relaxation was active.

PREs fall off with the sixth power of the distance, yielding a limited useful distance range. On the other hand, they offer the possibility to study minor conformational species of molecules and transient complexes in solution.[120,121]

Pseudocontact shifts (PCS)

Pseudocontact shifts (PCS) are defined as the difference in chemical shifts between the paramagnetic and diamagnetic states. The PCS gives information on the distance and angle between a nucleus and the paramagnetic center according to Equation 1.3:[113]

$$\Delta\delta^{PCS} = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}\left(3\cos^2\theta - 1\right) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\varphi\right] \qquad (1.3)$$

where r, θ, and φ are the polar coordinates of the nucleus with respect to the principle axes of the χ-tensor and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic components of the χ-tensor, respectively. The r^{-3} distance dependency is much weaker than the r^{-6} distance dependency of PRE. PCS can therefore be measured for nuclear spins that are far from the paramagnetic center, not strongly affected by PRE. Figure 1.3 depicts the characteristics of PRE and PCS.[90,111]
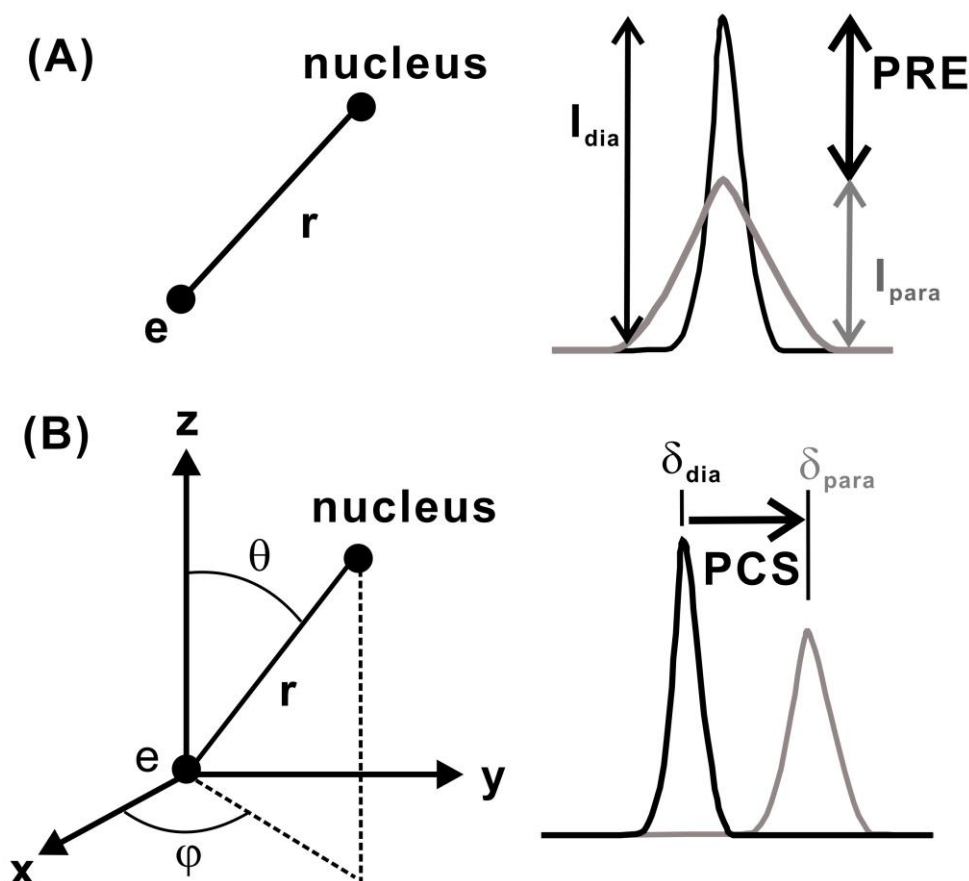
Figure 1.3: Principles of PRE and PCS. (A) PRE depends on the distance between electron and nuclear spin, and is observed as intensity difference between the paramagnetic and diamagnetic states. (B) PCS is observed as chemical shift differences between the paramagnetic and diamagnetic states. For PRE and PCS the intensity ratios or chemical shift differences of NMR resonances between a paramagnetic vs. a diamagnetic protein are used for structural studies.

1.5 Generation of paramagnetic effects on biomolecules

The use of paramagnetic effects in metalloproteins has been pioneered by Bertini *et.al.*[114,116] Later it was expanded from metalloproteins to non-metalloproteins by external paramagnetic tags. Common approaches to introduce paramagnetic effects on the target protein are presented:

Direct metal substitution

Lanthanides (Ln) do not exist in natural biological systems. For metalloproteins containing calcium, magnesium and manganese, the similarity between trivalent lanthanides (Ln^{3+}) and Ca^{2+} in terms of ionic radius and oxophilicity allows direct metal substitution, providing a valuable tool for structural dynamic studies.[122–127] Cu^{2+}- and Fe^{3+}-binding proteins were also used for obtaining paramagnetic restraints without substituting the metal ions.[128,129]

Metal-chelating tags

In recent years, a variety of site-directed lanthanide-binding tags (LBTs), spin-label compounds and lanthanide binding peptides (LBPs) have been developed to provide great opportunities for applying paramagnetic NMR on non-metallic proteins.

LBPs that mimic metal binding sites of metalloproteins can be inserted by standard molecular biology techniques, at either terminus or in the loop regions of non-metalloproteins, to serve as paramagnetic center once they are chelated to the appropriate metal.[130–135] This type of peptides can also be attached to the protein via cysteine modification.[136]

Another way to incorporate Ln is to use organic compounds as LBTs. The attachment of LBTs through cysteine thiol modifications is advantageous since it affords a rational means of generating Ln-tagged proteins with desired orientations between the Ln ion and the target proteins for NMR application.[133] However, site-specific incorporation of cysteines and removal of other solvent-accessible cysteines can sometimes results in problems in folding and purification. These problems can be eliminated by using metal-binding unnatural amino acids.[137] The amino acids can be incorporated site-specifically by a modified expression system.[138] Despite its relatively weaker binding affinity to lanthanides compared with other LBTs and LBPs, this approach has no limitation on the amino acid sequence, and therefore provides more possibilities in expanding NMR structural biology.[139]

Spin labels

Nitroxide spin labels are the simplest and most widely used paramagnetic tags for PRE.[89] These nitroxides contain an unpaired electron in a sterically protected environment, making them stable radicals under physiological conditions.[140] In most cases, they are usually attached to a surface cysteine residue, and therefore are mobile groups. This has to be taken into account when deriving PRE restraints.[141] Recently, an amino acid with a stable radical and a rigid conformation, 2,2,6,6-tetramethylpiperidine-1-oxyl-4-amino-4-carboxylic acid (TOAC) has been introduced into peptides by using conventional solid phase synthesis.[142] Due to its relatively bulky structure, placing TOAC outside the recognition motif is preferable to minimize the effect on binding and to extract PRE information.[143]

1.6   Thesis outline

The aim of this thesis is to study protein-small molecule and protein-peptide interactions by various NMR approaches, including NOE, PCS and PRE. Particular interests are emphasized on paramagnetic NMR in the following topics: (1) implementation of a general methodology of paramagnetic NMR to elucidate ligand binding modes; and (2) dynamics of transient protein-peptide encounter complexes.

Chapter 2 describes the initial ligand studies for FKBP12. In order to characterize the behavior of the ligands and the chemical shift perturbations on the protein target, a set of small-molecule ligands selected from a TINS competition screen and known FKBP12 ligands from literature were studied by 1D-^{1}H and [^{1}H, ^{15}N]-HSQC NMR. It was found that the chemical structure and the binding affinity of the ligand could have dramatic effects on their behavior in the 1D-^{1}H diamagnetic and paramagnetic NMR spectra.

Chapter 3 discusses the NOE-based structure of a small-molecule ligand bound to FKBP12. Multi-dimensional NMR was applied to complete the assignments of the protein in the free and bound form with the ligand. The ligand was first identified in TINS screening and later validated with chemical shift perturbations (CSPs). However, CSPs alone were not sufficient to elaborate the binding site, as perturbations were found in both site 1 and site 2 of FKBP12. Therefore, the actual binding site was determined by intermolecular NOE restraints from isotope filtered/edited NOESY experiments. Despite possible dynamics present in both the ligand and the protein, the site is in agreement with the binding sites identified for known FKBP12 ligands.

In Chapter 4, the same protein-small molecule structure was determined independently by a paramagnetic NMR approach, mainly using ligand PCS. The paramagnetic effect was implemented by using a lanthanide-binding tag, CLaNP-5.[144] The structural calculations were carried out in two individual ways: with either the predicted or the optimized tensors. A comparison between these two PCS-derived structures and the NOE-derived structure is presented. The advantage and disadvantage of this methodology compared with other NMR approaches is addressed.

In Chapter 5, the interactions between tetralysine peptides and plastocyanins from different species – poplar tree (*Populus nigra*), Japanese fern (*Dryopteris crassirhizoma*), and a cyanobacterium (*Phormidium laminosum*) – were investigated by PRE NMR and theoretical Monte Carlo simulations. Tetralysine peptides have been shown to interrupt the binding between Pc and cytochrome *f* and to cause subtle structural changes on the copper site of Pc.[44,145–147] In this study, lysine peptides that contain TOAC spin-labels[143] were applied to investigate the dynamics of the transient Pc-peptide complexes.

Finally, in the concluding remarks in Chapter 6, the results of these studies are summarized and integrated.

2

NMR characterization of the binding behavior of FKBP12 ligands

Abstract

This chapter describes a pilot study of qualitative characterization of ligand binding to FKBP12 using 1D-^{1}H NMR in order to reveal their suitability for a follow-on paramagnetic NMR study. We include examples of difficulties that can be encountered during the process of structural characterizations for hits identified from ^{1}H-NMR based screening. A variety of FKBP12-binding small molecules, selected from the literature or identified from a TINS screen, were characterized for their binding to FKBP12. As a result of combination of their binding affinities, chemical structures and differences in relaxation between the para- and diamagnetic states, these small molecules show different behavior in the NMR spectra between the free and bound, as well as the para- and diamagnetic forms. Through this analysis, we have established the requirements for ligands suitable for structure determination of the ligand-protein complex using paramagnetic NMR spectroscopy. The results may serve as guidelines for choosing suitable ligands and optimizing experimental conditions for paramagnetic NMR studies.

2.1   Introduction

NMR spectroscopy has been widely used in characterizing structures of organic molecules. Because of the high gyromagnetic ratio and abundance of protons present in organic molecules, many fragment-based screening techniques have been developed based on analyzing 1D-^{1}H spectra of the compounds screened.[68,148–150] These screening techniques, however, are mostly applicable for primary screening but not for structural determination of the ligand-protein complex. Some of them have been developed to provide limited information on ligand binding modes.[77]

Paramagnetic NMR spectroscopy has been applied in studying protein-protein interactions, but rarely used in protein-ligand interactions. It is useful to use paramagnetic NMR due to its wide range of choices for the paramagnetic source and predictable long-distance effects on the systems of interest.[89,111,117] In order to extend the utility of proton chemical shifts of the ligands to elaborate detailed protein-ligand interactions by using paramagnetic NMR spectroscopy, it is required to examine the behavior of the ligands in ^{1}H-NMR upon binding to a protein. To date, there have been three cases studying protein-ligand structure exclusively with paramagnetic NMR restraints.[96,97,99] Therefore, the application of paramagnetic NMR in protein-ligand studies may still be in its infancy.

In this study, we focused mainly on simple 1D-^{1}H NMR measurements of the protein-ligand complexes, either with or without the paramagnetic source. The 12 kDa FK-506 binding protein (FKBP12), a well-characterized protein with many known small-molecule ligands, was chosen as the model protein.[30] To suppress the resonances of the protein in the spectra, a T_2 relaxation filter was applied using the Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence. For the paramagnetic NMR study, a lanthanide-binding tag CLaNP-5 (Caged Lanthanide NMR Probe 5)[144] was attached to the protein to introduce paramagnetic effects to the protein-ligand complex. In total, fifteen ligands with diverse chemical structures were selected: three ligands reported in structure-activity relationship (SAR) studies,[30,151] four ligands from docking studies[152] of FKBP12, and eight ligands were FKBP12 hits identified from a TINS (Target-immobilized NMR screening)[84] screen.

Here we focus on two subjects frequently discussed in NMR: line broadening and chemical shift changes. First, tumbling time increases when the molecular size increases. Small molecules usually have sharp lines due to their fast tumbling time. When the ligand is bound to the protein, it behaves more similar to the protein. Therefore the NMR resonances of the ligand are broadened in the bound form. With the T_2 relaxation filter, the intensity of the

resonances from the species that have shorter relaxation time, such as the protein and the bound ligand, would be dramatically reduced, resulting in a cleaner spectrum dominated by the species which have longer relaxation time, such as the free ligand. This simple concept has been applied in many NMR screening techniques, including TINS.[84]

Second, ligand binding to its target protein induces chemical shift changes in both the protein and ligand with varying magnitudes that are dependent on the local chemical environment. Chemical environment of the ligand nuclei changes upon binding, which is reflected from the corresponding changes in the chemical shifts of the ligand nuclei. In the fast exchange regime in terms of NMR time scale, the observed chemical shifts are the weighted average of chemical shifts from the free and bound ligand. For slow-exchange species, the bound and free populations show separate signals. For intermediate-exchange species, the resonances can be very broad and sometimes difficult to observe in the spectra.

The positions and linewidths of the ligand ^{1}H resonances were qualitatively examined in 1D-^{1}H NMR spectra, in the presence of either the wild type or, in some cases, CLaNP-5 tagged FKBP12. The ligands were found to exhibit varied or unchanged positions and linewidths in 1D ^{1}H-NMR spectra, not only between their free and bound form, but also between the para- and diamagnetic environment. Possible factors that could contribute to the different behavior of the ligands were discussed. Finally, we conclude the requirements for a suitable ligand for the paramagnetic NMR study.

2.2 Materials and Methods

Ligand preparation

The known FKBP12 ligands from SAR studies−ZB6, ZB10 and ZB88−were synthesized according to the synthetic schemes described in the literature.[30,151] The other ligands were purchased from the suppliers listed in Table 2.1.

The ligands were dissolved in H_2O buffer, D_2O buffer or d_6-ethanol with a final concentration 25-100 mM. Resonance assignments of the ligands were achieved by acquiring 2D [^{1}H, ^{13}C]-HSQC and [^{1}H,^{13}C]-HMBC spectra acquired on a 100 mM solution of the ligand in D_2O or d_6-DMSO.

Table 2.1: Details of supplier information for the ligands used in this study.

ZB code	Supplier	Catalog number	CAS
ZB1	Sigma-Aldrich	155101	1137-68-4
ZB2	Sigma-Aldrich	642711	1137-67-3
ZB3	Sigma-Aldrich	226769	670-96-2
ZB84	Sigma-Aldrich	D87759	91-44-1
ZB293	Maybridge	S13756	--
ZB390	InterBioScreen	STOCK1N-10205	--
ZB429	Sigma-Aldrich	54169-9	--
ZB1035	Labotest	LT00367766	--
ZB1051	Chemical Block	A1180/0054672	--
ZB1104	Sigma-Aldrich	S292249	--
ZB1406	InterBioScreen	STOCK1N-60294	--
ZB1489	Maybridge	JFD02405	--

Protein expression and purification

Recombinant human FKBP12 wild type and the double cysteine variant K34C/K35C, all of which contain an additional LEHHHHHH tag at the C-terminus were purified from *Escherichia coli* strain BL21(DE3) containing the overexpression plasmid pET20b with FKBP12 insert. The bacterial culture was incubated at 37 °C until OD 0.6-0.8, and then gene expression was induced by adding IPTG to a final concentration 1 mM. Incubation was continued at 18 °C overnight. After centrifugation, the pellets were resuspended in 5 mL/g pellet sucrose buffer (50 mM Tris, 20% sucrose, 1 mM EDTA, pH 7.5), centrifuged at 7,000 × g for 30 min at 4 °C, re-suspended in 5 mL/g pellet of 5 mM $MgSO_4$ and incubated on ice for 10 min. Cells were then centrifuged at 4,500 × g for 20 min, the supernatant was discarded and the pellet was resuspended in 1.5 mL/g pellet lysis buffer (50 mM sodium phosphate, 5 mM imidazole, 500 mM NaCl, pH 7.5) and frozen. For lysis, cell suspensions were incubated with PMSF, DNase and lysozyme at room temperature for 1 h and lysis was achieved using a French Press (SLM Instruments Inc.). The crude lysate was cleared by ultracentrifugation at 35,000 × g for 45 min at 4 °C followed by filtration through a 0.22 μm syringe filter. FKBP12 was purified to homogeneity using a 5 mL His-Trap column (GE Healthcare) with a gradient of 5-500 mM imidazole, followed by a Superdex G75 gel filtration column (GE Healthcare). Protein concentrations were determined by UV-Vis

spectroscopy ($\varepsilon_{280} = 9970$ M^{-1}cm^{-1}). The purity was confirmed by SDS-PAGE and Coomassie blue staining. The yield was in general 80 mgL^{-1} for wild type FKBP12 and 15 mgL^{-1} for the double cysteine mutants.

CLaNP-5 attachment

To a solution of FKBP12 double cysteine mutant in 25 mM sodium phosphate, 100 mM NaCl, pH 7.5, 10 mM dithiothreitol was added and the reaction mixture was incubated on ice for 1 h to reduce cystines. After removing DTT with a PD-10 column (GE Healthcare), 3-10 molar equivalents of Ln^{3+}-CLaNP-5 were added to a 20 μM solution of the reduced FKBP12 and incubated on ice for 1 h. The mixture was purified using a PD-10 column followed by a 24 mL Superdex 75 gel filtration column (GE Healthcare) to separate the probe attached monomeric protein from dimers and excess probe. Protein concentrations were determined by UV-Vis spectroscopy ($\varepsilon_{280} = 9970$ M^{-1}cm^{-1} for FKBP12 and 2200 M^{-1}cm^{-1} for CLaNP-5). The fractions of dimers were estimated to be 10%-25% based on the intensities observed on SDS-PAGE analysis. The yields of pure tagged proteins after purification were 20%-40%.
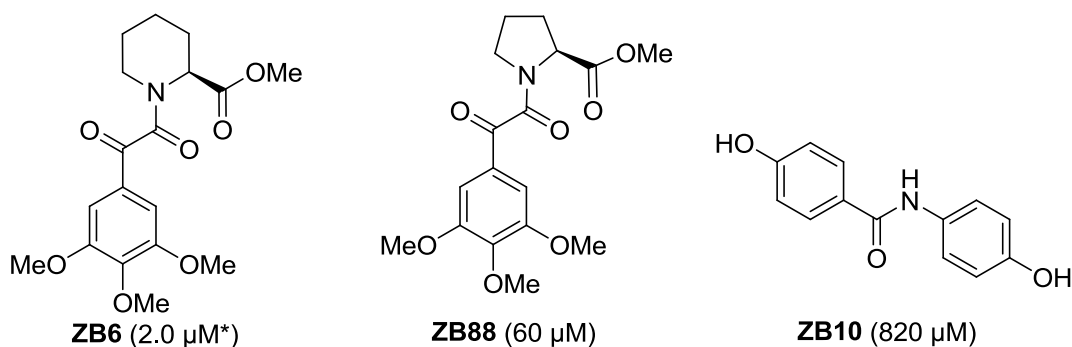
NMR measurements

All protein NMR samples contained 15 mM Tris-HCl, 25 mM NaCl, pH 7.7 and >95% D$_2$O. The concentration of wild type FKBP12 and Ln-CLaNP-5 attached FKBP12 was 10 to 100 μM for titrations. Spectra were recorded at 290 K on a Bruker Avance DMX-600 spectrometer equipped with a TCI-Z-GRAD cryoprobe. 1D-^{1}H spectra of the protein-ligand complexes were recorded at 600 MHz with a spectral width of 16 ppm and 6 k complex points, resulting in a digital resolution of 1.56 Hz before zero filling. The length of the transverse relaxation filter was semi-optimized based on the visibility of the resonances of each free ligand, and therefore varied per ligand, ranging from 0.5-200 ms. Data were processed and analyzed in TopSpin (Bruker).

2.3   Results

To find clear paramagnetic pseudocontact shifts (PCSs) of the ligands, we aimed to look for ligands that have clear, measurable changes on ^{1}H chemical shifts upon binding to their target protein. FKBP12, a well-characterized 12-kDa protein with many known small-molecule ligands, was selected as the model. In order to assess the effect of ligand binding on 1D-^{1}H NMR spectra of the protein-ligand complexes, fifteen compounds were selected for the initial tests. Figure 2.1 shows an overview of the compounds studied in this chapter. The first group consists of three ligands selected from SAR studies,[30,151] in which ZB6 and ZB88 are FKBP12 site-1 ligands and ZB10 a site-2 ligand. The second group consists of four

FKBP12 ligands selected from a docking study.[152] Two common characteristics are present in ZB1, ZB2 and ZB3: an arylimidazole skeleton which contains only aromatic protons (apart from the imido protons), and similar affinity for FKBP12. ZB84 is structurally different from the other three compounds in this group. The third group includes nine compounds identified as FKBP12 hits from an NMR screen of a collection of commercially available fragment-like molecules. The collection shows a large variation in structures, some of which have only aliphatic protons and some others have both aromatic and aliphatic protons.

FKBP12 ligands from SAR studies

ZB6 (2.0 μM*)   **ZB88** (60 μM)   **ZB10** (820 μM)

FKBP12 ligands from docking studies

ZB1 (83 uM)   **ZB2** (63 uM)   **ZB3** (41 uM)   **ZB84 (**130 uM*)
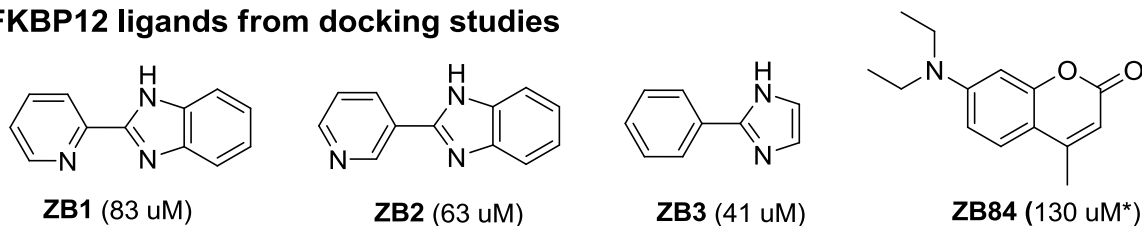
FKBP12 ligands from TINS screening

ZB293 (1.8 mM)   **ZB390** (3.6 mM)   **ZB429** (2.4 mM)   **ZB1035** (0.5 mM)

ZB1051   **ZB1104**   **ZB1406**   **ZB1489**
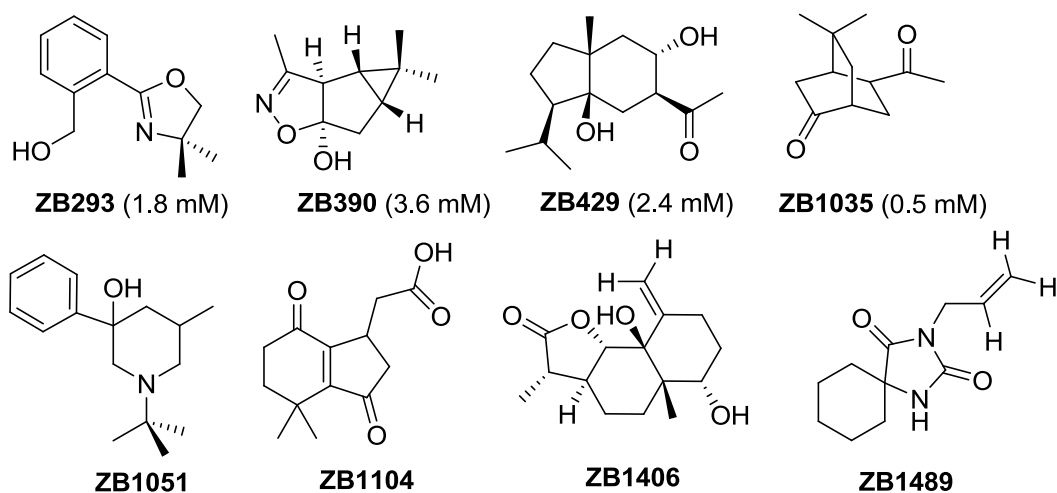
Figure 2.1: Overview of FKBP12-binding compounds discussed in this chapter. The available K_D values are shown in brackets. K_D values of ZB1, ZB2, ZB3, ZB293 and ZB1035 were determined by NMR titrations. K_D values of ZB1051, ZB1104, ZB1406 and ZB1489 were not determined. *K_D values for ZB6[30] and ZB84[152] were taken from references.

Presence of isomers, slow to intermediate exchange, and scalar coupling

Our first attempt was to determine the structure of a known complex of FKBP12 and a bound ligand using paramagnetic NMR, and to compare with the structure of the same complex reported in literature. ZB6 was chosen as the starting point for our initial study because of the available SAR by NMR information for the complex of FKBP12 and ZB6, exhibiting a K_D of 2 μM.[30] Characterization of the ligand binding started with the titration of ZB6 against wt FKBP12. However, the proton resonances of ZB6 did not show clear chemical shift changes or line-broadening upon binding to FKBP12 (Figure 2.2). The only resonances that showed slight line-broadening during the titration were the resonances belonging to the methoxy groups (Figure 2.2, protons 1, 2 and 3). Besides, careful inspection of the spectra of free ZB6 suggested that at least two isoforms of ZB6 were present in the solution. Neither of the two forms showed clear shifts during the titration. Therefore, ZB6 may already exhibit slow exchange on the NMR time scale. The resonances of the bound ligand were not found in the spectra due to the use of the relaxation filter.
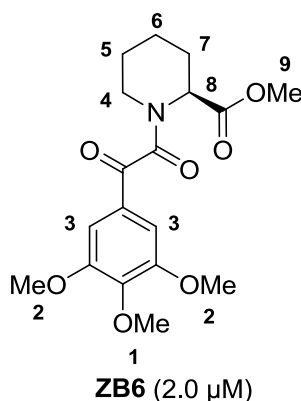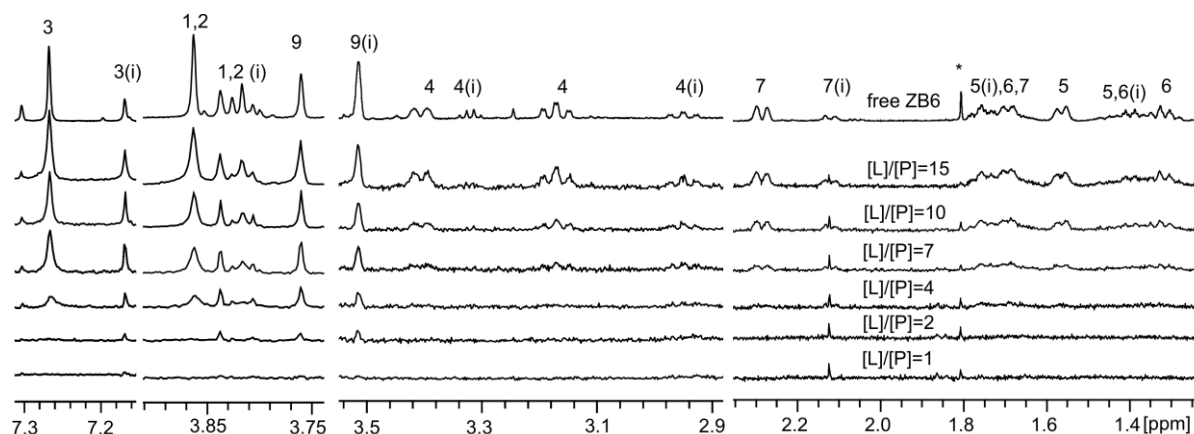
Figure 2.2: 1D-^{1}H NMR spectra of ZB6 titrated into wt FKBP12. The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). Asterisks indicate the protein/buffer resonances. Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. The concentration of the protein was 10 μM and free ZB6 150 μM. The fraction of bound ligand was 6.5% at the end point of the titration. The relaxation filter was 60 ms.

We then set out to assess the binding behavior of ZB88, an analog of ZB6, using CLaNP5-tagged FKBP12. CLaNP5 is a two-armed lanthanide binding tag which can be attached to the protein via disulfide bond linkage.[144] The cysteines required for CLaNP5 tag attachment

26

were introduced at residues 34 and 35, resulting in a K34C/K35C variant of FKBP12. Thulium (Tm) was used as the paramagnetic lanthanide, and lutetium (Lu) as the diamagnetic control in CLaNP5. ZB88 has been shown to bind to FKBP12 in a computational SAR study.[151] In a follow-on study, the [^{1}H,^{15}N]-HSQC spectra of FKBP12 showed visible shifts upon binding of ZB88 with K_D 60 µM (data not shown).

When ZB88 was added to the CLaNP5-tagged FKBP12, the proton resonances of ZB88 showed very different linewidths in the presence of paramagnetic protein compared to the diamagnetic protein. First, in the 1D-^{1}H NMR spectra of free ZB88, two forms were identified (Figure 2.3, top panel), similar to the situation of ZB6. Second, the major form of ZB88 binds FKBP12 and experiences severe paramagnetic line broadening, resulting in only one resonance of ZB88 indentified from the paramagnetic spectra (H-8), whereas the rest were broadened beyond detection. A possible reason for this is that binding of ZB88 must be in the intermediate exchange regime on the NMR time scale. The resonances of the minor isoform showed little to no paramagnetic shifts during the course of titration, in which only
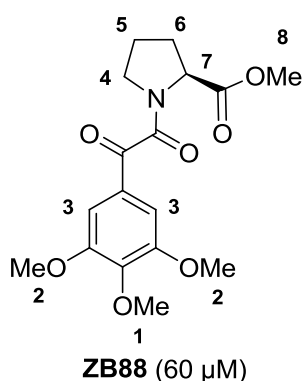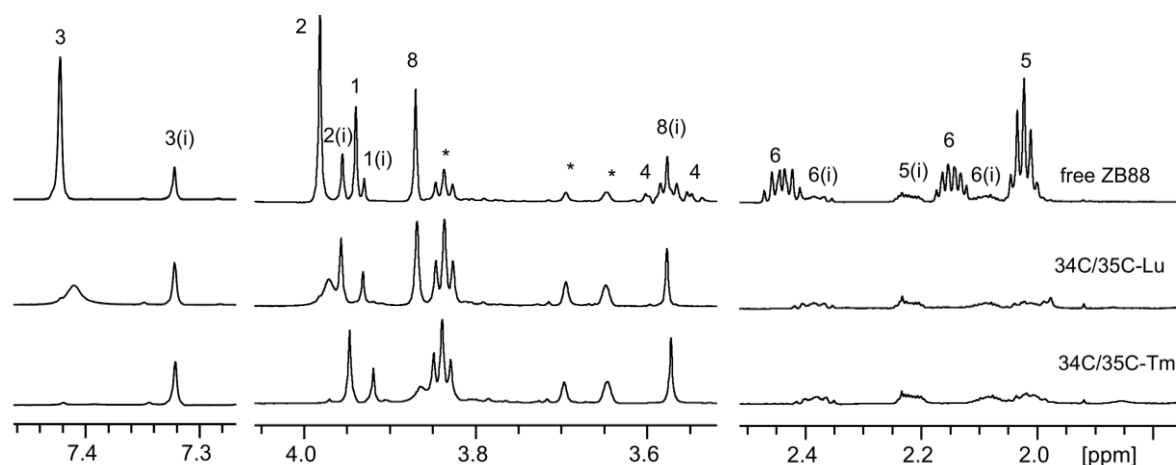
Figure 2.3 : 1D-^{1}H NMR spectra of ZB88 in the free form (top trace), and in the presence of diamagnetic FKBP12 (34/35-Lu, middle trace) and paramagnetic FKBP12 (34/35-Tm, bottom trace). The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. Asterisks indicate the protein/buffer resonances. [FKBP12]=75 µM, [ZB88]=2000 µM, resulting in a ligand : protein ratio of 26.7 and a bound ligand fraction of 3.6%. The T_2 filter was 50 ms.

H-1(i) and H-2(i) showed shifts and the others no shifts. All three visible resonances are methyl groups, showing the highest peak intensity among other protons. Third, the intensities of the resonances which have strong scalar couplings, such as the protons of the proline ring, were dramatically reduced in the spectra. Given only one ligand PCS was available from the major isoforms, severe line-broadening due to intermediate exchange, and the presence of isomers, it was not practical to proceed with structural calculations.

The two species of ZB88 were thought to be two of the three possible conformational isomers in equilibrium (Figure 2.4). The major species is thought to be the conformation with the *trans* form for both the two ketone groups and the proline group (Figure 2.4, conformational isomer 1) because it is more energetically favored. It seems not possible to isolate isomers 1 and 3. Isomer 2 can appear due to the *cis-trans* prolyl isomerase function of FKBP. The combination of intermediate exchange, complex scalar coupling and ligand isomers present in equilibrium in solution has made it difficult to use this ligand for the paramagnetic study. Due to its limited solubility in water, the ligand concentration could not be higher.
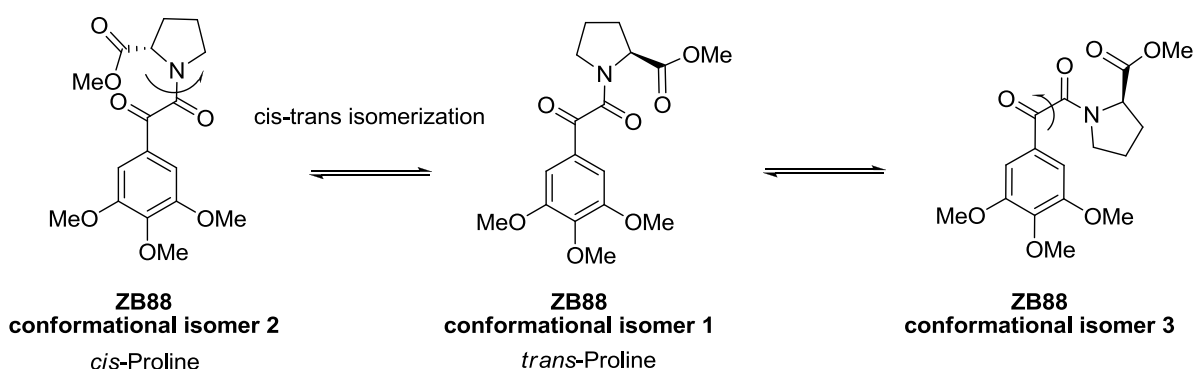
Figure 2.4: Proposed conformational equilibrium in ZB88.

Symmetric structure

We then tested another FKBP12 ligand, ZB10, that is known to bind at the so-called 'site 2' of FKBP12. The ligand was added to the CLaNP5-attached FKBP12 at a protein to ligand ratio of 1 to 1. The binding affinity of ZB10 against wild type FKBP12 (K_D 0.8 mM) was determined by observing gradual chemical shift changes of the amides of FKBP12 using [^{1}H, ^{15}N]-HSQC in the presence of saturating concentration of ZB88 (data not shown). Eight protons of ZB10 are visible in the NMR spectrum and appear as four doublets due to degeneracy of NMR signals resulting from the symmetric structure (Figure 2.5, top trace). Therefore, only four ligand PCSs could be obtained (Figure 2.5, middle and bottom traces).

Each PCS corresponds to the average of the degenerate proton pair. The available ligand PCS data can still be used for structural calculations, but the number of available restraints would be reduced due to degeneracy.
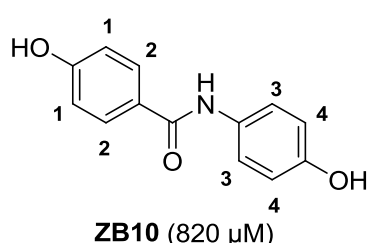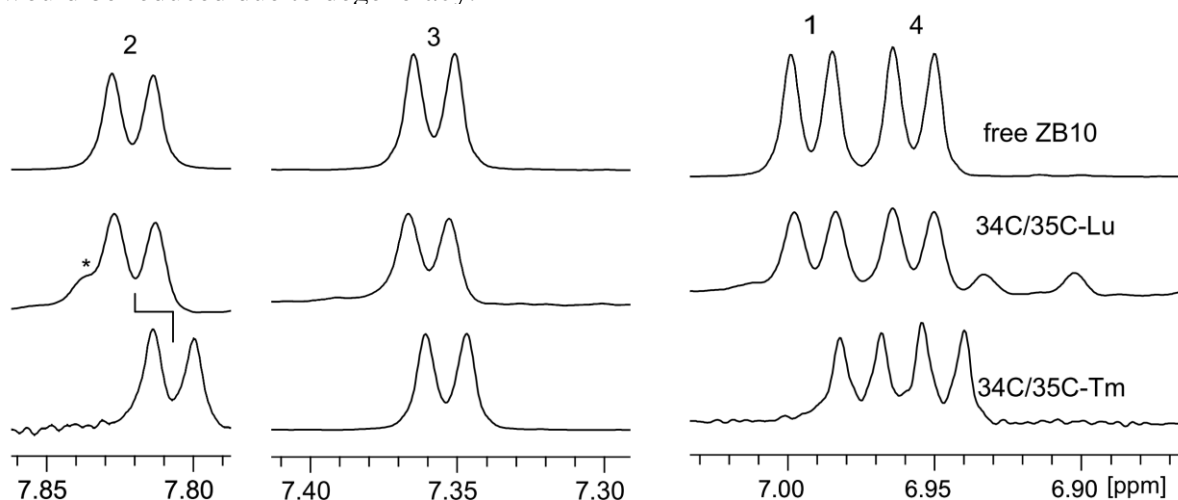
Figure 2.5: 1D-^{1}H NMR spectra of ZB10 free form and in the presence of FKBP12 (K34C/K35C) tagged with CLaNP-5 (Lu, Tm). The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). The PCSs are indicated by connecting lines. Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. Asterisks indicate the protein resonances. The protein and ligand concentrations were 23 μM and 30 μM, resulting in 2.7% of bound ligand and 3.5% of bound protein. The T_2 filter length was 200 ms.

In a previous study, ZB10 was identified as the site 2 ligand of FKBP12 in the presence of saturating concentration of ZB6.[30] Binding of the same ligands have not been tested in the opposite way (saturate with ZB10 and then add ZB6). Here, ZB10 was added into the CLaNP5-tagged FKBP12 before the addition of ZB88. Surprisingly, ZB88 does not bind FKBP12 when ZB10 was already present (Figure 2.6). It should be noted that in this condition FKBP12 was not yet saturated by ZB10 (27% of FKBP12 and 2% of ZB10 were bound, based on K_D of 0.8 mM). The PCSs from ZB10 were still visible upon addition of ZB88, while the chemical shifts of the resonances of ZB88 in the para- and diamagnetic samples appeared to be identical. This indicated that ZB10 is not only a FKBP12 site 2 binder, but also a site 1 binder and therefore competes with ZB88 for the same binding site. Thus the K_D of ZB10 determined in the presence of ZB88 (0.8 mM) might not be applicable in this situation.
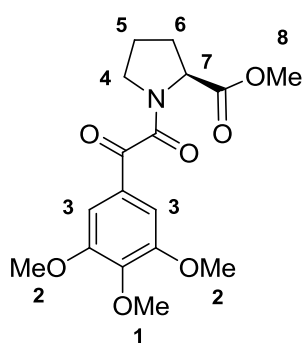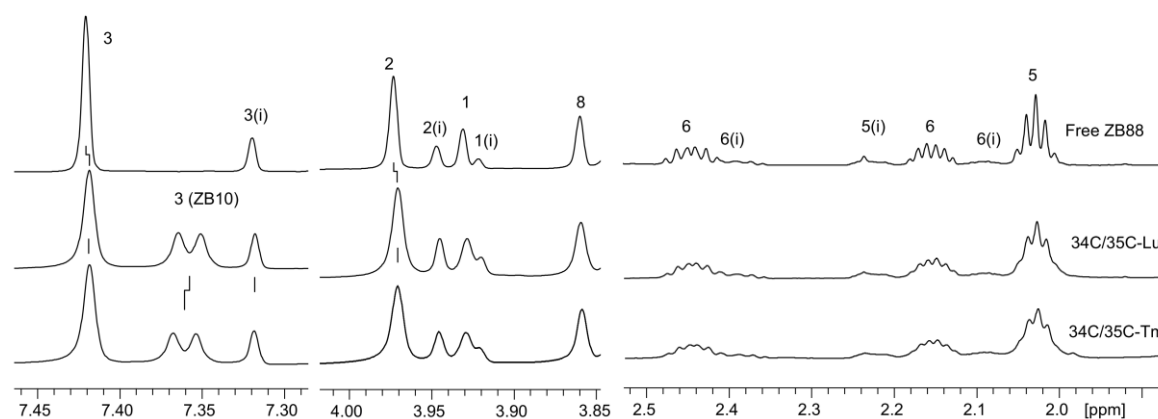
Figure 2.6: 1D-^{1}H NMR spectra of ZB88 titrated into wt FKBP12 in the presence of ZB10. The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). The label "3(ZB10)" represents the protons 3 from ZB10. Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. Resonances labeled with (i) represent the protons from ZB88 isomers. The concentration of FKBP12 and ZB10 was 23 μM and 300 μM, resulting in 27% bound protein and 2.1% bound ligand. No PCS could be observed for ZB88 in the concentration range from 150-600 μM. The spectra above contain 600 μM ZB88. The T_2 filter length was 100 ms.

Symmetric structure and intermediate exchange

ZB1, ZB2, ZB3, ZB84 were ligands selected from a docking study on FKBP12.[152] The common feature of ZB1, ZB2 and ZB3 is the arylimidazole scaffold. The binding affinities and chemical shifts of the resonances in ^{1}H NMR spectra are also similar. ZB84 contains a coumarin scaffold and its binding affinity is different from the other three ligands in the group. Four out of six of the proton resonances of ZB2 can be identified in the spectra although the linewidths were very broad due to aggregation of the ligand at high concentration (Figure 2.7). This was also observed in the broader linewidths at the ligand to protein ratio of 20 than at the ratio of 10. The severe line broadening in some resonances could be due to intermediate exchange, as in the case of ZB88. The spectra of ZB1 (Figure A1, three out of six resonances were broadened beyond detection) and ZB3 (Figure A2, one resonance was broadened beyond detection and one resonance was very broad) are provided in Appendix A.
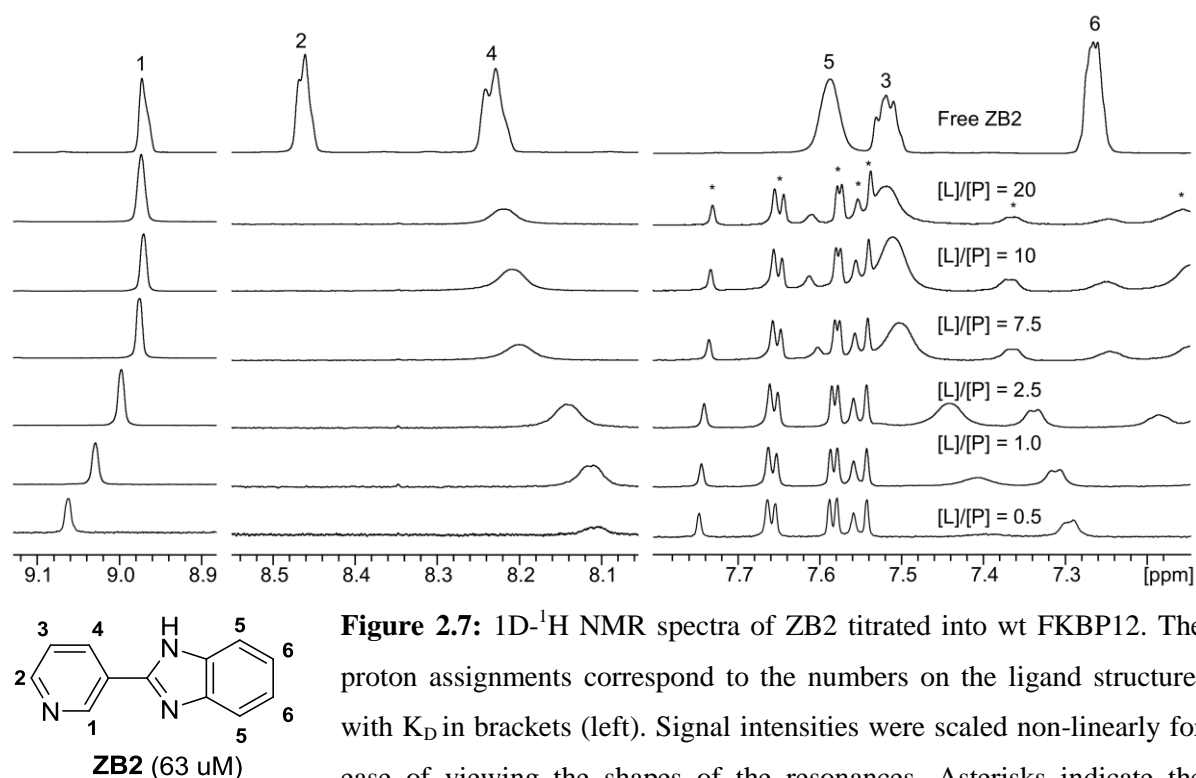
Figure 2.7: 1D-^{1}H NMR spectra of ZB2 titrated into wt FKBP12. The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. Asterisks indicate the protein resonances. The initial concentration of the protein was 100 μM and the concentration of free ZB2 was 2 mM. The fraction of bound ligand was 4.8% at the end point of the titration. The T_2 filter length was 50 ms. The broad resonances in the free ligand spectra was caused by aggregation of ZB2 at a concentration higher than 1 mM, regardless of the presence of FKBP12.

On the contrary, the resonances of ZB84 were completely broadened beyond detection (Figure 2.8). Two resonances (protons 4 and 7) were likely to be visible, but unfortunately due to peak overlapping with the buffer (proton 7) and the protein (proton 4) resonances, none of the resonances were clearly distinguishable. It could be that the relaxation filter was too long (50 ms) in this case, or the bound fraction of the ligand (1.9%) needs to be reduced. This example also points out the limitation of 1D ^{1}H-NMR due to resonance overlap.
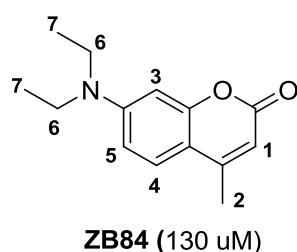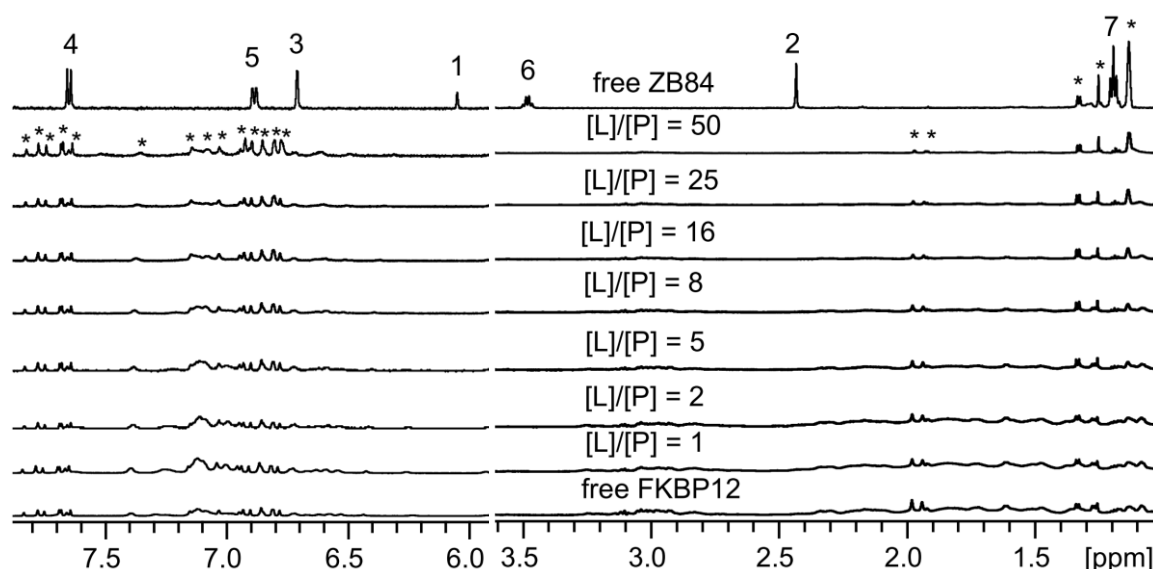
Figure 2.8: 1D-^{1}H NMR spectra of ZB84 titrated into wt FKBP12. The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. Asterisks indicate the protein and solvent resonances. Concentration of free FKBP12 and free ZB84 was 50 μM and 2.5 mM. The bound ligand fraction was 1.9% at the end point of the titration. The T_2 filter length was 50 ms.

Table 2.2 shows a summary of the above mentioned four ligands studied with wild type FKBP12. Although these ligands have similar K_D values as ZB88, they do not exhibit complicated scalar coupling. Some of the resonances of ligand were still broadened beyond detection in their bound form, indicating intermediate exchange.

Table 2.2: Summary of FKBP12 ligands selected from docking studies.

Ligand	ZB1 (83 uM)	ZB2 (63 uM)	ZB3 (41 uM)	ZB84 (130 uM)[a]
Protons	9	9	8	17
Resonances (**free ligand**)	6	6	4	7
strong, sharp (***)	1	1	2	0
strong, broad (**)	2	2	0	0
weak, broad (*)	2	2	1	1
beyond detection (−)	1	1	1	6

(Resonances (bound ligand) label spans the bottom four rows.)

[a] Data taken from reference[152].

Ligands selected from FKBP12 TINS screen

As none of the above mentioned ligands displayed ideal characteristics, we decided to investigate the hits identified from TINS screen. In total eight hits were tested. Two of them are shown as examples below, in which one ligand proved to be very suitable while another ligand exhibits challenges different from the previously discussed. The spectra of the other ligands are provided in Appendix A, Figures A3-A8.

ZB293 has both aromatic and aliphatic protons and a well-resolved 1D-^{1}H NMR spectrum with simple scalar couplings. The ligand was tested with FKBP12 and the chemical shift changes of the resonances of the ligand upon binding to FKBP12 were clearly visible (Figure 2.10), indicating fast exchange in NMR timescale. Besides, the structure of ZB293 is asymmetrical, which does not contribute to NMR degeneracy. The only degeneracy is the methyl groups in the free form, which were resolved in the protein bound state. This can be due to the fixed conformation upon binding.
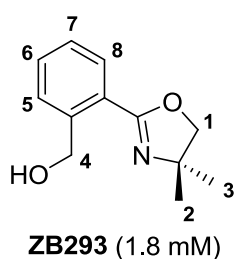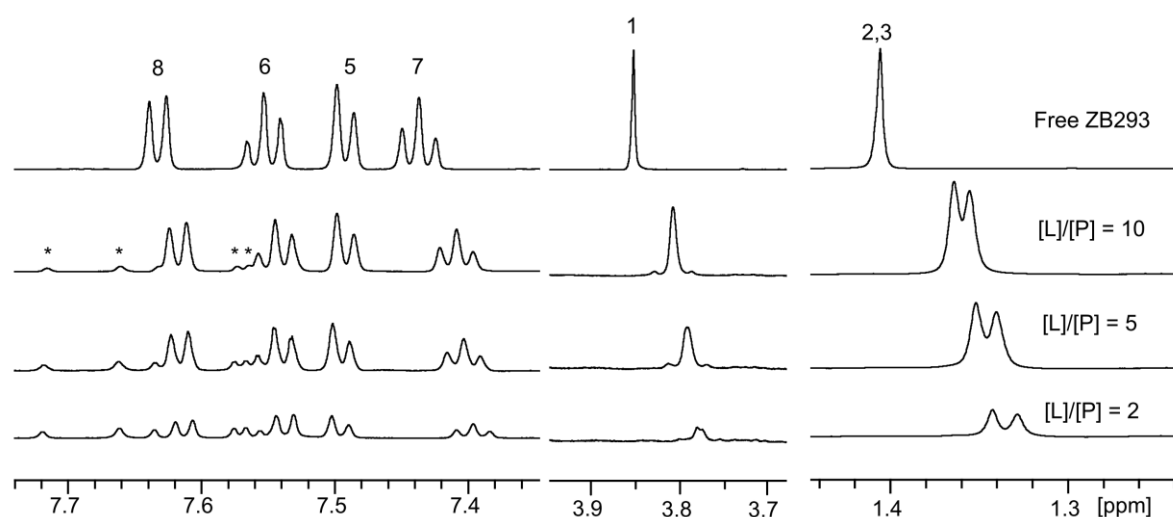
Figure 2.10: 1D-^{1}H NMR spectra of ZB293 titrated into wt FKBP12. The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. Asterisks indicate the protein resonances. Free protein and free ligand concentration was 100 and 1000 μM. The fraction of bound ligand was 3.5% at the end point of the titration. The T_2 filter length was 40 ms.

NMR-identical isomers

In synthesis procedures, drugs having one or more centers of chirality were often obtained as isomeric mixtures of both diastereomers and/or enantiomers. Below is an example encountered during our study related to this issue.

ZB1035 was an FKBP12 hit identified from TINS screen whose binding was verified by analysis of chemical shift perturbations from [^{1}H,^{15}N]-HSQC spectra of FKBP12 upon titration of the compound. Although the compound has two chiral carbons, it showed promisingly high ligand efficiency (binding energy/non-hydrogen atoms,[56] 0.314 kcal/mol) and was potentially interesting as a starting point for lead optimization. However, the chemical shifts of the ligand itself were not examined closely. When free in solution, ZB1035 shows a single form in the ^{1}H-NMR spectrum (Figure 2.11, top panel). When the ligand is bound to FKBP12, most of the ligand resonances split into 2 sets of resonances. Both sets of resonances showed two separate sets of chemical shifts that gradually merged into one single set at increasing ligand concentration. This observation indicates that (1) there are two isoforms present in the ligand, and (2) it is possible for both forms of the ligand to bind FKBP, but the individual affinities can be difficult to determine by simple NMR experiments. The population of these two species in the compound mixture is unknown. The source of these different forms can be from the synthesis procedure.
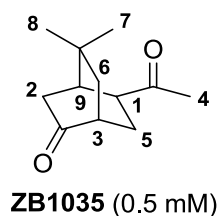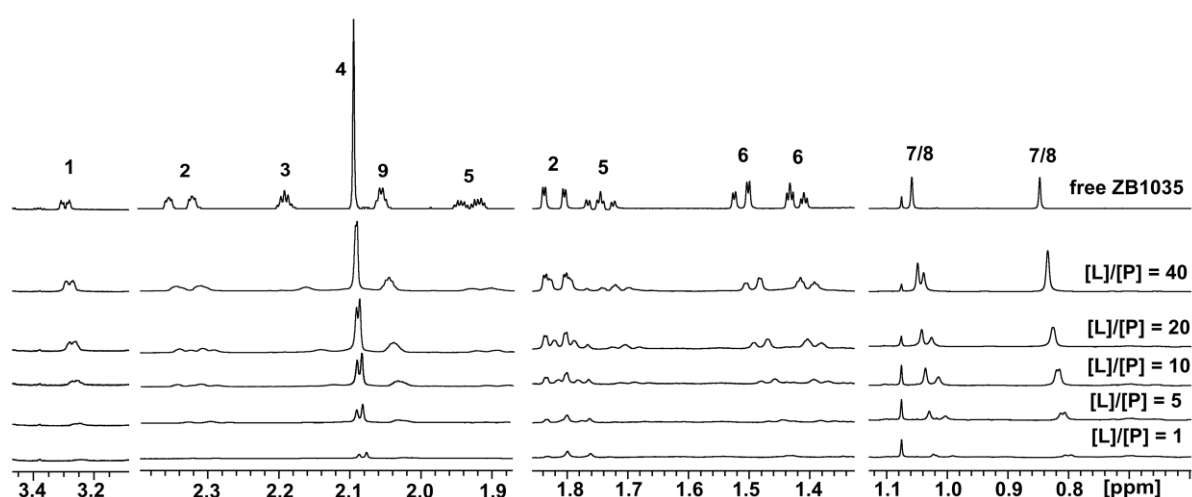
Figure 2.11: 1D-^{1}H NMR spectra of ZB1035 titrated into wt FKBP12. The proton assignments correspond to the numbers on the ligand structure, with K_D in brackets (left). Many resonances in the bound form have split into two. For example, protons 2, 4, 7 and 8 consist of two sets of peaks, which gradually merge into one with increasing ligand concentration. Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. The unlabeled resonance at 1.08 ppm is from the buffer. The initial protein concentration and free ZB1035 concentration was 48 μM and 1920 μM, resulting in a total of 2.0% bound ligand at the end point of the titration. The T_2 filter length was 20 ms.

Inspection of the structure of ZB1035 indicates that this compound can have even four different isomers in the free form (Figure 2.12), including enantiomers which are NMR-identical, and diastereomers which might be distinguishable given a correct structural model. By calculating the coupling constant between H^9 and H^1 (2.65 Hz) in the NMR spectra and comparing with the models obtained from PRODRG server[153] (Figure 2.12B), it is likely that isomers **a** and **b** are present in the mixture. Although this analysis reduces the number of isomers from four to two, no information concerning the ratio of **a** and **b** is available. As this situation is complicated and there was not enough of the compound for further structural analysis, the study of this compound was discontinued. To obtain an enantiomerically pure compound, a chiral separation of the racemic mixture can be used, or a chiral paramagnetic reagent[154] can be used to determine the ratio of the enantiomers.
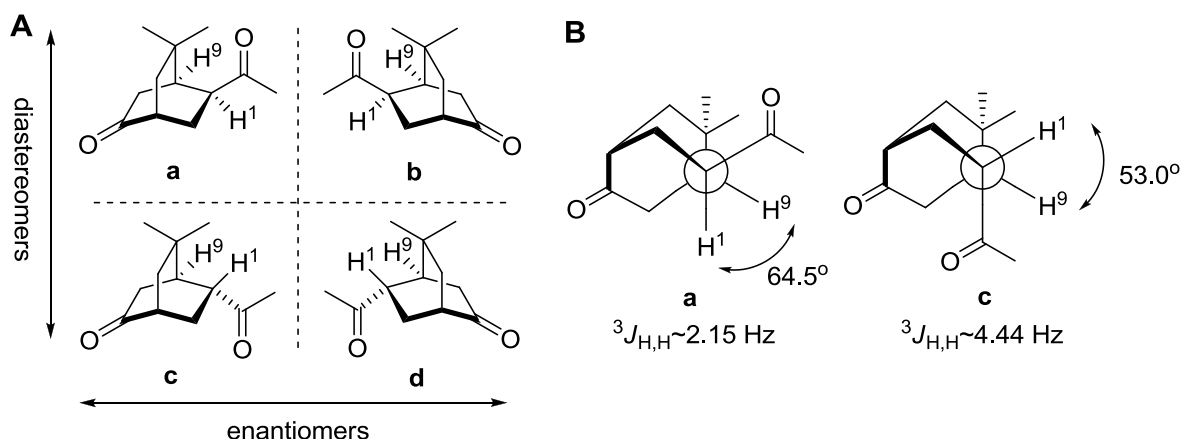
Figure 2.12: (A) Possible isomers that are present in ZB1035. From left to right are the enantiomers. From top to bottom are the diastereomers. (B) Newman projections of ZB1035 isomers a and c. Models used for calculation of dihedral angles were generated from PRODRG server.[153] Coupling constants calculated according to the empirical generalization of Karplus equation.[155]

2.4 Discussion

Although three paramagnetic NMR studies on protein-ligand complexes using 1D ^{1}H-NMR have been published,[96,97,99] limited practical information concerning the applicability and limitations is available and therefore this area still requires further development.

We have performed a series of studies to characterize the effect of ligand binding to FKBP12 on the 1D-^{1}H NMR spectrum of the ligand. The ultimate goal was to observe clear changes in ligand chemical shifts from simple 1D-^{1}H NMR spectra upon binding to the target protein. To achieve this, it was concluded that the requirements for a suitable ligand to study are: (1) has an asymmetric structure, (2) contains well-resolved proton resonances with minimal degenerated protons, (3) no stereoisomers or enantiomers present in the compound, (4) the

scalar couplings shall not obscure assigning the resonances in the dia- and paramagnetic spectra, (5) binding at the regime of fast exchange in NMR time scale, and (6) has clear changes in chemical shifts upon binding to its target. ZB293 satisfies all the above requirements and therefore is the best candidate for further paramagnetic NMR studies among all the ligands investigated here. In Chapter 3 and 4 the interaction of ZB293 with FKBP12 will be studied in detail.

Structurally symmetric fragments, such as ZB10, are still possible to be studied, but would present some difficulties due to reduced number of restraints. Ligands that show slow exchange are possible to study if the resonances of the bound ligand can be identified and assigned from the spectra.[99] Ligands that show severe line broadening due to scalar coupling or intermediate exchange, such as ZB88 and ZB84, would be difficult to measure ligand PCSs. Ligands that contain isomers, such ZB6, ZB88 and ZB1035, should be best isolated to a single form prior to such NMR studies.

Besides the consideration on the ligand structure, there are also other requirements to optimize the set-up of 1D-^{1}H NMR experiments:

1. The length of the relaxation filter should be long enough to eliminate most of the protein resonances while reducing as little as possible the intensity of the ligand resonances. The relaxation filter length used here is between 0.5-200 ms. In theory, it is possible to optimize the length of T_2 filter by simulating the relaxation decay of the protein and the ligand. However, it proved to be difficult as every proton on the ligand has different relaxation time and every ligand exhibits different affinity to the target. Therefore, the range of T_2 filter length provided here is only a guideline. The optimal T_2 filter length still can be protein- and ligand-dependent.

2. Avoid peak overlap with the resonances from the buffer or solvent. Most of the protein signals have been dramatically reduced because of the relaxation filter and the use of deuterated buffer conditions. Nevertheless, there can still be some residual signals, particularly in the aliphatic region.

3. Sufficient fractions of bound ligand in fast exchange are required. It is essential to maintain a small fraction bound of a ligand in fast exchange in order to observe sharp lines (with preferably low protein concentration). This varies for each compound, depending on the chemical structure and binding affinity.

A new NMR methodology, called pure-shift NMR,[156–158] may be able to address the problem of multiple, strong J couplings. This technique eliminates multiplets and increases the resolution of ^{1}H-NMR, although the sensitivity might be sacrificed. For compounds with

symmetric structures, a combination of ^{13}C-, ^{19}F-, or ^{31}P-NMR with ^{1}H-NMR can increase the number of observable chemical shift changes in the resonances between the free and the bound form from 1D NMR spectra.

3

Structure determination of a protein-ligand complex by NOE

Abstract

This chapter presents the determination of ligand binding pose by the traditional NOE approach. Chemical shift perturbations (CSPs) were used to determine K_D and initially also to determine the binding site. However, CSPs monitor both direct and indirect effects, as observed in the complex. Therefore isotope filtered/edited NOESY measurements were used to determine the ligand binding mode. To characterize the interaction, full assignments on the protein and the ligand were performed. Many intermolecular NOEs were identified from the NOESY spectra and subsequently applied for structural calculation. The resulting structure is used as a reference structure to compare with the structure of the same complex calculated based on paramagnetic NMR data (Chapter 4).

3.1 Introduction

Structure information of protein-ligand complexes is highly valuable in the early stages of structure-based drug design (SBDD) and fragment-based drug discovery (FBDD), where X-ray crystallography plays an important role to provide structure information. In many cases, however, hit compounds that interact weakly with their targets may not readily crystallize, due to a variety of reasons. Due to its high sensitivity in detecting weak interactions, NMR has proved to be a very useful tool when X-ray crystallography cannot be applied.

Ligand-induced chemical shift perturbations (CSPs) are commonly used to monitor ligand binding and, for rigid proteins, to determine ligand binding sites. However, CSPs caused by direct contact with ligand and indirect conformational changes cannot be distinguished by simply mapping CSPs on the protein structure.

NOEs develop due to through-space interactions rather than through-bond interactions. The intensity of an NOE is proportional to the inverse of the sixth power of the distance that separates the two dipolar-coupled spins within 5 Å, and so NOEs are sensitive probes of short-range through-space intramolecular and intermolecular interactions. In standard isotope-edited NOESY spectra, intermolecular NOEs are indistinguishable from intramolecular protein NOEs. Acquiring 2D and 3D edited/filtered NOESY[159,160] on a sample of $^{13}C/^{15}N$-labeled protein saturated with unlabeled ligand can overcome this problem. This type of experiments detects only NOEs between protons directly bound to ^{13}C or ^{15}N (protein) and protons bound to ^{12}C or ^{14}N atoms (ligand) while suppressing all other cross-peaks by isotope filtering and editing.[161] Crosspeaks from the filtered NOESY spectrum can therefore be unambiguously assigned to protons on the small-molecule ligand.

The aim of the work presented here is to characterize the structure of a small molecule ligand bound to a 12 kDa FK-506 binding protein (FKBP12) using CSP and intermolecular NOE restraints. FKBP12 is a peptidyl-prolyl isomerase which belongs to the family of immunophilins and is a drug target for the immuno-suppressants rapamycin and FK506. There have been many structural studies on this protein.[9,10,15,20,162,163] The ligand in this study is a fragment that was identified as a hit against FKBP12 from a screen of a fragment library using Target-Immobilized NMR Screening (TINS).[84] Using high-resolution NMR spectroscopy, we determined the binding mode of the bound ligand. The determined ligand binding mode serves as a reference structure for Chapter 4, in which a paramagnetic NMR approach was applied to determine the binding mode of the same ligand.

3.2   Materials and Methods

Ligand preparation

The ligand **1**, [2-(4,4-dimethyl-5,5-dihydro-1,3-oxazol-2-yl)phenyl]methanol, was purchased from MayBridge (catalog number S13756). Assignment of the ligand resonances was achieved from 2D ^{1}H, ^{13}C-HSQC and ^{1}H,^{13}C-HMBC spectra acquired on a 100 mM solution of the ligand in D_2O.

Protein expression and purification

Uniformly ^{15}N-labeled and ^{15}N, ^{13}C-labeled FKBP12 was purified from *Escherichia coli* cells grown on minimal media containing ^{15}NH$_4$Cl and ^{12}C$_6$- or ^{13}C$_6$-glucose, respectively. The proteins were essentially expressed and purified as described in Chapter 2.

NMR measurements

All protein NMR samples contained 15 mM Tris-HCl, 25 mM NaCl, pH 7.7, 6% D_2O for ^{15}N-labeled protein and >95% D_2O for non-isotope-labeled protein. The concentration of wild type FKBP12 was 100 μM for titrations, and 1.5 mM for resonance assignment and NOESY experiments. 1D-^{1}H, [^{1}H,^{15}N]-HSQC, HNCA, HNCACB, HN(CO)CA, HN(CO)CACB, HNCO, (H)CCH-TOCSY, HN(CA)CO and HBHA(CBCACO)NH spectra were recorded at 290 K on a Bruker Avance DMX-600 spectrometer equipped with a TCI-Z-GRAD cryoprobe. A ligand to protein ratio of 10:1 was used in the 3D ^{15}N-separated ω1-^{13}C/^{15}N-filtered NOESY-TROSY, 3D NOESY-[^{1}H,^{15}N]-TROSY, 3D ^{13}C-separated ω1-^{13}C/^{15}N-filtered NOESY-[^{1}H,^{13}C]-HSQC, 3D NOESY-[^{1}H,^{13}C]-HSQC, 2D ^{13}C-edited, ^{13}C/^{15}N-filtered NOESY, and a ratio of 3:1 used in 2D NOESY, 2D DQF-COSY, and 2D TOCSY spectra of the complex were recorded on a Bruker Avance 950 MHz spectrometer with a ^{1}H{^{13}C,^{15}N} cryogenic probe. Data were processed in TopSpin (Bruker) and then spectra were analyzed in Sparky.[164] Dr. Frank Löhr (Goethe University Frankfurt, Germany) is acknowledged for the NMR measurements on the 950 MHz spectrometer.

Calculations of dissociation constants and bound ligand fractions

Ligand binding was observed via the changes of protein resonances in the [^{1}H,^{15}N]-HSQC spectrum upon titration with the ligand.[165] For analysis of the chemical shift perturbations (CSPs) of ^{1}H and ^{15}N backbone resonances, the weighted average chemical shift values were calculated and normalized according to equation 3.1:

$$\Delta\delta_{avg} = \sqrt{\frac{1}{2}[(\frac{\Delta\delta_N}{5})^2 + \Delta\delta_H^2]} \qquad (3.1)$$

where $\Delta\delta_N$ and $\Delta\delta_H$ are the differences of ^{15}N and ^{1}H chemical shift of an amide group, respectively.

The dissociation constant (K_D) was determined using a two-parameter non-linear regression curve fitting based on a one-site binding model as described in equation 3.2:

$$\Delta\delta_{avg} = \frac{1}{2}\Delta\delta_0(A - \sqrt{A^2 - 4R})$$

$$A = 1 + \frac{1}{R} + \frac{P_0 R + L_0}{P_0 L_0\left(\dfrac{1}{K_D}\right)} \qquad (3.2)$$

where R is the total [ligand] to [protein] ratio, $\Delta\delta_{avg}$ is the average CSP (equation 3.1) at a given R, $\Delta\delta_0$ is the CSP at 100% bound protein, P_0 is the starting concentration of the protein, L_0 is the stock concentration of ligand and K_D is the dissociation constant.[166] The fraction of bound ligand was calculated using the dissociation constant.

NOE-based structure calculations

Intermolecular NOE crosspeaks were identified in 2D and 3D NOESY spectra with NOE mixing times of 50-70 ms and then converted into distances using the CYANA[167] calibration function. Intermolecular restraints were introduced as ambiguous restraints if degenerate protons were involved or the protein resonance could not be unambiguously assigned.

Protein coordinates were taken from the X-ray and NMR structures of FKBP12 (PDB entries 1FKR,[9] 1FKS,[9] 1FKT,[9] 1D6O[27] and 2PPN[29]) and the backbone was kept fixed throughout the process. Starting structures for the complex were generated by placing the ligand in random orientations with respect to FKBP12. Then the NOE distance restraints were applied to generate ligand orientations which satisfied the intermolecular restraints. Side chain atoms within 8 Å of the ligand were allowed to rotate during subsequent energy minimization. The XPLOR-NIH script is available in Appendix B.

3.3 Results and Discussion

Characterization of the ligand-protein interaction

The ligand **1** (Figure 3.1A) was identified from a TINS screen[84] of a library of commercially available, low molecular weight "drug fragments" for binding to FKBP12. To confirm the binding of **1** to FKBP12 and obtain structural insight into the binding site, we titrated the

ligand into ^{15}N labeled protein and observed CSPs in a series of [^{1}H,^{15}N]-HSQC spectra with increasing ligand concentration. Figure 3.1B shows the titration curves for the ten residues most effected from which the equilibrium dissociation constant, K_D, could be extracted. CSPs occurred throughout the protein, including in two previously defined ligand binding sites, referred to as site-1 and site-2.[30] Five of these residues are site-1 residues and the other five residues belong to site-2. In Figure 3.1C the CSPs have been mapped onto the crystal structure of FKBP12 (PDB entry 2PPN[29]) and color coded according to their magnitude. While CSPs can be caused by direct changes in the electronic environment of spins close to the ligand, many other factors can contribute. For example, it has been shown that, upon ligand binding to FKBP12, perturbations of both main-chain and side-chain dynamics can occur at sites distal to the binding interface.[168] Since such changes in dynamic behavior may also lead to CSPs, it was not possible to define the ligand binding site by CSP mapping alone. Therefore we sought an alternative method to elucidate the structure of the complex.
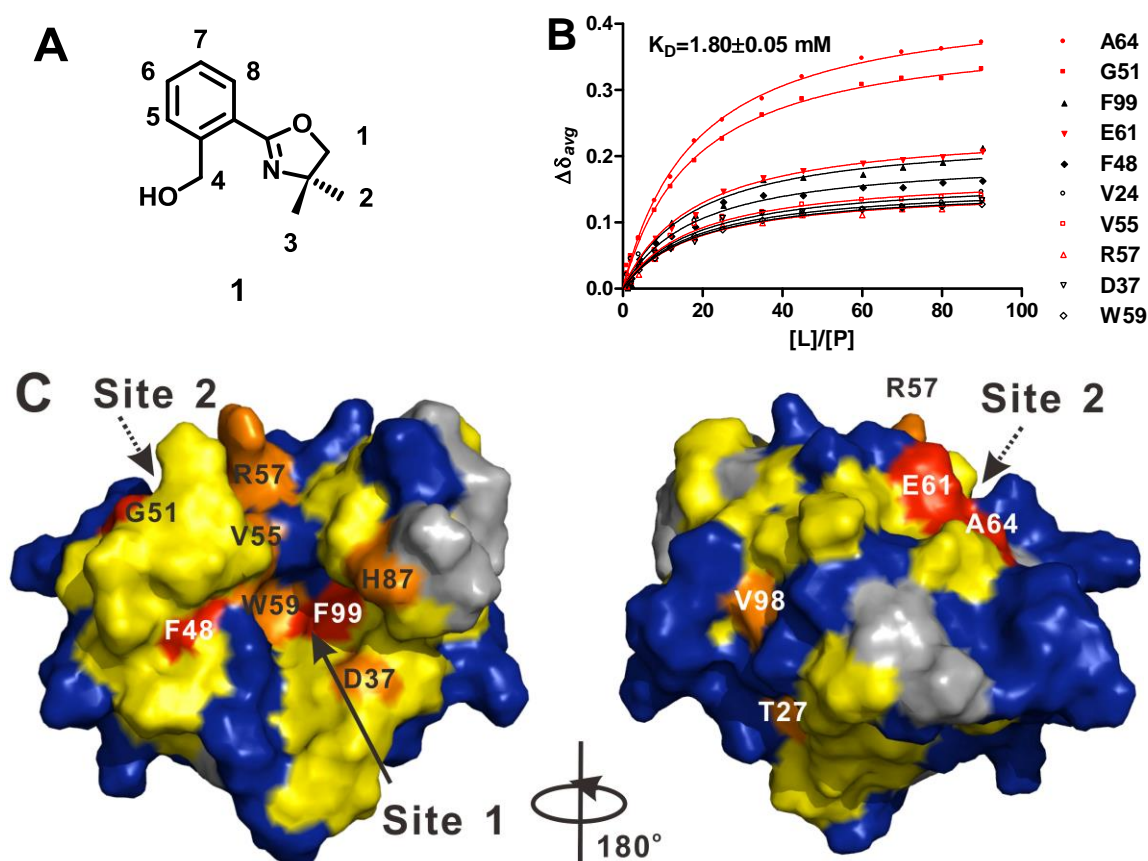
Figure 3.1: (A) Chemical structure of ligand 1. (B) Chemical shift changes of FKBP12 resonances as a function of increasing [L]/[P]. The top 10 residues which showed largest perturbations are shown. Residues in the site-1 and site-2 regions are shown in black and red, respectively. The dissociation constant of 1 was obtained by fitting simultaneously to a 1:1 binding model (equation 3.2, solid lines). (C) Mapping of CSPs from the binding of 1 on the structure of FKBP12 (PDB accession number

2PPN[29]). The positions of site 1 and site 2 are indicated. Color representation: red, $\Delta\delta_{avg} > 0.15$ ppm; orange, $0.15 > \Delta\delta_{avg} > 0.10$ ppm; yellow, $0.10 > \Delta\delta_{avg} > 0.03$ ppm; blue, $\Delta\delta_{avg} \leq 0.03$ ppm; dark grey, no data. Figures showing protein structures were prepared with PyMOL.[169]

NMR assignments

Assignment of the chemical shifts of free FKBP12[7] and FKBP12 bound to ascomycin[12] have been reported and these were used as the starting points. The complete assignments were verified as following: Backbone assignments (HN, CO, C_α and C_β) of ligand-free protein were obtained from standard triple resonance spectra and transferred to the bound state by following changes in [^{1}H,^{15}N]-HSQC and [^{1}H,^{13}C]-HSQC spectra upon ligand addition. Subsequently, the aliphatic side chains were assigned from a (H)CCH-TOCSY spectrum and the aromatic ones from 2D ^{1}H,^{1}H-NOESY, DQF-COSY and ^{1}H,^{1}H-TOCSY spectra.

NOESY measurements

NOEs develop due to through-space dipole-dipole interactions, thus containing information on the distances between atoms separated by 5Å or less in space. The intensity of an NOE is proportional to the inverse of the sixth power of the distance between the two nuclei, but not always a precise reflection of the distance. The structure of the FKBP12-**1** complex was initially determined using standard intermolecular NOE based methods. Intermolecular NOEs were derived from a combination of 3D ^{15}N-separated ω1-^{13}C/^{15}N-filtered NOESY-TROSY, 3D NOESY-[^{1}H,^{15}N]-TROSY, 3D ^{13}C-separated, ω1-^{13}C/^{15}N-filtered NOESY-[^{1}H,^{13}C]-HSQC, 3D NOESY-[^{1}H,^{13}C]-HSQC, 2D ^{13}C-edited,^{13}C/^{15}N-filtered NOESY and 2D NOESY spectra. Figure 3.2 and 3.3 show regions of NOESY spectra containing intermolecular NOEs between the ligand and backbone amides (Figure 3.2) and side chains (Figure 3.3) of the protein. In total, 66 intermolecular NOE crosspeaks were identified in these NMR spectra.
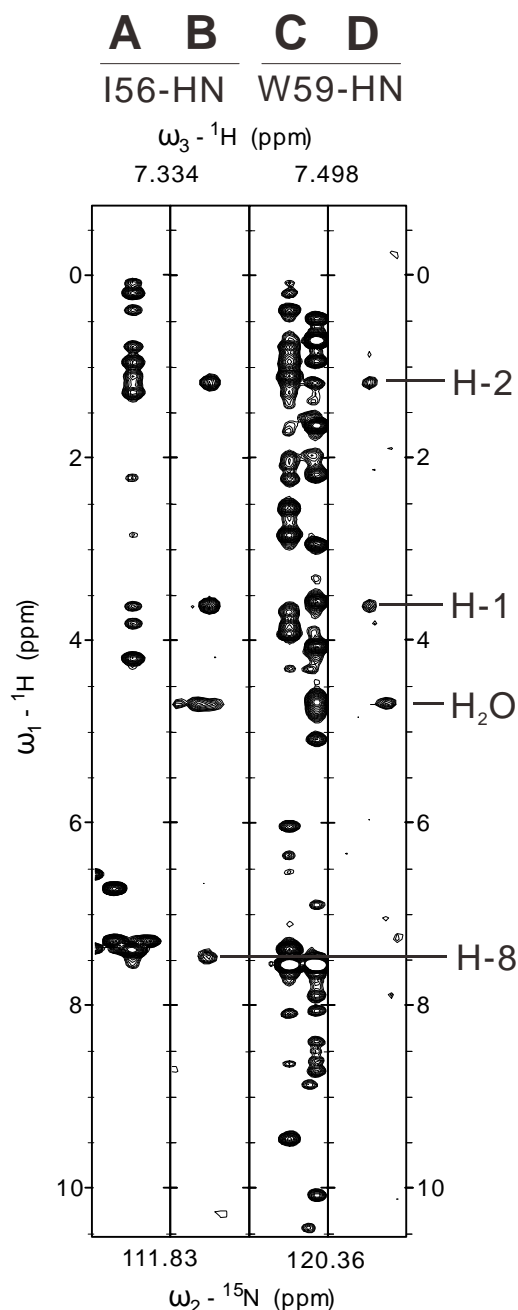
Figure 3.2: ^{1}H-^{1}H cross-sections from NOESY-^{15}N TROSY (A, C) and ^{13}C,^{15}N- filtered NOESY-^{15}N-TROSY (B, D) of FKBP12-1 complex, displaying intermolecular NOE cross-peaks between selected backbone amide groups of ^{13}C/^{15}N labeled FKBP12 and ^{1}H nuclei of the unlabeled ligand 1. (A, B) I56 at (δ_N,δ_H) = (111.83, 7.334) ppm and (C, D) W59 at (δ_N,δ_H) = (120.36, 7.498) ppm. Chemical shifts of ligand protons and H_2O are indicated.

Ligand conformation

There are two possible orientations of the five-membered oxazole ring relative to the six-membered aromatic ring. Between these two orientations, the five-membered ring flips by 180°. The difference in intensities of intramolecular NOEs between protons 4/1, 4/2 and 4/3 (proton numbers as indicated on the structure in Figure 3.1A) in the presence of FKBP12 suggests that the presented conformation is the prefered bound conformation in solution. A possible explanation is that the hydroxyl group from the aromatic ring may form a hydrogen bond with the nitrogen atom, creating an extra 7-membered ring.[170]
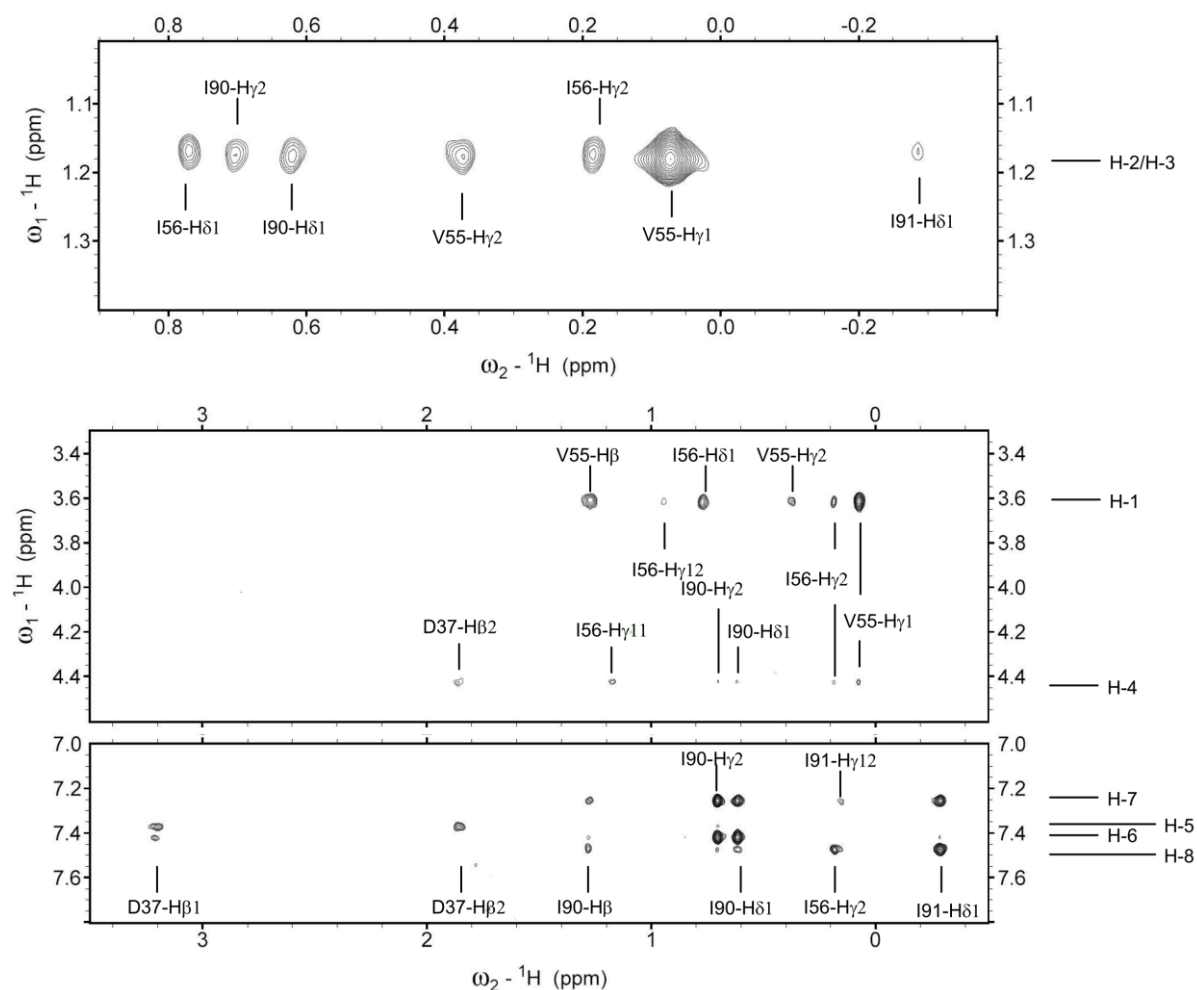
Figure 3.3: Regions showing intermolecular NOE evidences in 2D ^{13}C-edited,^{13}C/^{15}N-filtered NOESY. Chemical shifts of the ligand ^{1}H nuclei are indicated on the right.

NOE-based structure calculations

Among the 66 intermolecular NOE crosspeaks, 43 restraints were used for structural calculations (Table 3.1). The NOEs that were weak (20 peaks) or with uncertain assignments (3 peaks) were excluded. Superposition of various FKBP12 structures from the PDB (1FKR, 1FKS, 1FKT, 1D6O, 2PPN) indicated that the loops surrounding the hydrophobic pocket are variable and hence might undergo significant dynamic behavior in solution. Therefore, to cover the range of conformations of the loops, multiple structures were used as input for structure calculations.

Table 3.1: List of intermolecular NOE restraints and their corresponding distances (in Å) calculated from the final NOE structure. (A) Restraints from flexible residues. (B) Restraints from rigid residues. (C) Restraints that were excluded due to weak intensities or uncertain assignments.

(A) NOE restraints from flexible residues

no.	residue #	protein	ligand	r(exp)	r(calc)
1	37	Hβ1	H-5	3.4	4.1
2	37	Hβ2	H-5	3.4	3.3
3	55	Hα	H-1	3.4	2.6
4	55	Hα	H-2,H-3	3	3.6
5	55	Hα	H-8	3.4	6.5
6	55	Hδ	H-1	3.3	4.5
7	55	Hγ1	H-1	2.8	3.7
8	55	Hγ1	H-2,H-3	2.4	3
9	55	Hγ2	H-2,H-3	3.3	5.3
10	56	Hδ1	H-1	3.3	3.8
11	56	Hδ1	H-2,H-3	3.3	3.5
12	56	Hγ2	H-1	3.5	2.3
13	56	Hγ2	H-2,H-3	3.3	4.7
14	56	Hγ2	H-8	3.2	5
15	56	HN	H-1	2.8	2.4
16	56	HN	H-2,H-3	3	3.5
17	56	HN	H-8	3.4	6.4
18	82	Hδ	H-8	3.5	4.3
19	82	Hε	H-8	2.5	2.7
20	87	Hδ2	H-7	3.4	5.3
21	90	Hδ1	H-2,H-3	3.4	9.7
22	90	Hδ1	H-7	2.9	4.4
23	90	Hδ1	H-6	2.7	2.5
24	90	Hγ2	H-2,H-3	3.5	9.8
25	90	Hγ2	H-7	2.6	3.7
26	90	Hγ2	H-6	2.7	2.4
27	91	Hδ1	H-7	2.9	3.4
28	91	Hδ1	H-8	2.8	4.5

(B) NOE restraints from rigid residues

no.	residue #	protein	ligand	r(exp)	r(calc)
29	26	Hε	H-4	2.7	5.1
30	36	Hδ	H-5	2.9	3.3
31	36	Hε	H-5	3.5	3.9
32	46	Hδ	H-2	3.5	4.9
33	46	Hε	H-4	3.1	3.2
34	46	Hε	H-2	2.6	3.4
35	59	Hη2	H-3	3.4	3.7
36	59	Hη2	H-2	3.7	6.4
37	59	HN	H-1	3.5	6.4
38	59	HN	H-2,H-3	3.5	6.3
39	59	Hζ2	H-3	3.7	3.2
40	59	Hζ3	H-3	3.4	4
41	59	Hζ3	H-2	3.7	6.4
42	99	Hε	H-4	3.3	5.8
43	99	Hε	H-2	3.4	6.2

(C) Excluded NOE restraints

no.	residue #	protein	ligand	r(exp)	remark
44	37	Hβ1	H-4	3.8	weak
45	55	Hγ1	H-4	3.9	weak
46	55	Hγ2	H-1a	4.2	weak
47	55	HN	H-2,H-3	4.1	weak
48	56	Hγ11	H-4	3.5	weak
49	56	Hγ12	H-1	3.6	weak
50	56	Hγ2	H-1a	3.9	weak
51	56	Hγ2	H-1b	3.9	weak
52	56	Hγ2	H-4	3.9	weak
53	57	HN	H-2,H-3	3.8	weak
54	59	Hζ2	H-2	3.9	weak
55	60	HN	H-2,H-3	3.8	weak
56	90	Hβ	H-7	3.7	weak
57	90	Hβ	H-8	4.1	weak
58	90	Hδ1	H-4	3.6	weak
59	90	Hγ11	H-7	4.2	weak
60	90	Hγ2	H-4	3.8	weak
61	91	Hδ1	H-2,H-3	3.4	weak
62	91	Hγ11	H-7	3.9	weak
63	91	Hγ12	H-7	3.7	weak
64	82	Hε	H-1a	3.4	uncertain
65	99	Hδ	H-4	3.7	uncertain
66	99	Hζ	H-1a	3.4	uncertain

The position and orientation of the ligand is similar in all structures (overlay and individuals in Figure 3.4), suggesting that the conformation of the loops does not strongly influence the protein-ligand interaction. 33 out of 43 input NOEs were satisfied, resulting in a well-defined ligand orientation.
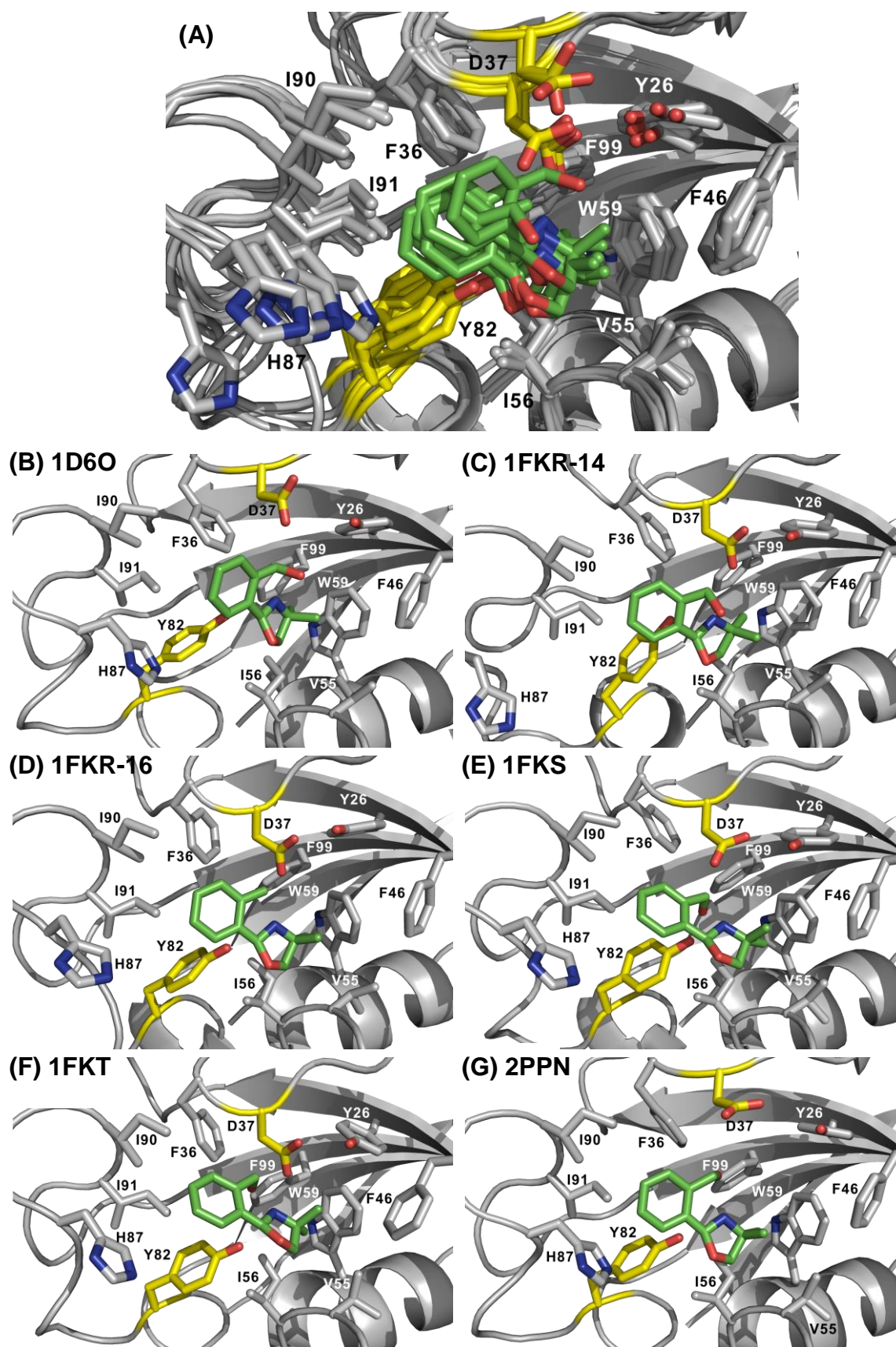
Figure 3.4: NOE-based structures calculated with selected PDB files. The side chains showing possible hydrogen bonding to ligand 1 are colored in yellow. The side chains of other residues showing intermolecular NOEs to the ligand are colored grey. (A) Overlay of the lowest energy

structures of the complex of **1** (in green sticks) with 6 previously determined structures of FKBP12 (1D6O, 1FKR models 14 and 16, 1FKS, 1FKT and 2PPN) as determined by intermolecular NOE restraints. The RMSD of all ligand atoms relative to the mean is 1.1 ± 0.4 Å. Each individual structure is presented in (B)-(G): (B) 1D6O (C) 1FKR model 14 (D) 1FKR model 16 (E) 1FKS (F) 1FKT (G) 2PPN.

However, some large violations of NOEs remain (Table 3.1 and Figure 3.5). The NOEs indicate **1** has contacts with residues located in site 1. The ligand interacts with the hydrophobic pocket formed by F36, F46, V55, I56, W59, I90, I91 and F99. The violations could be due to the motion of the flexible loop at residues 50-56 and 78-95 which are on opposite sides of the hydrophobic pocket.[163,168,171–173] As no single ligand orientation can satisfy all the restraints, it is possible that the dynamic behavior of the protein gives rise to time-averaged NOEs. Consistent with this idea, the backbone[163] and side chains[173] of FKBP12 were also shown previously to undergo chemical exchange on the μs-ms timescale. Alternatively, it is also possible that multiple ligand orientations are present. Therefore, the NOE structure presented here is only an approximation of the actual ligand binding mode. Analysis of the structures indicates possible hydrogen bonds present between (1) the ligand hydroxyl and Y82-hydroxyl, (2) the ligand hydroxyl and D37-Oδ and (3) the ligand nitrogen and Y82-hydroxyl. Previous studies have suggested that the side chains of D37[20,25] and Y82[15] are involved in hydrogen bonding with rapamycin and FK-506, two high affinity ligands of FKBP12. Our NOE-based structure model shares a similar hydrogen bonding framework with the previous studies.
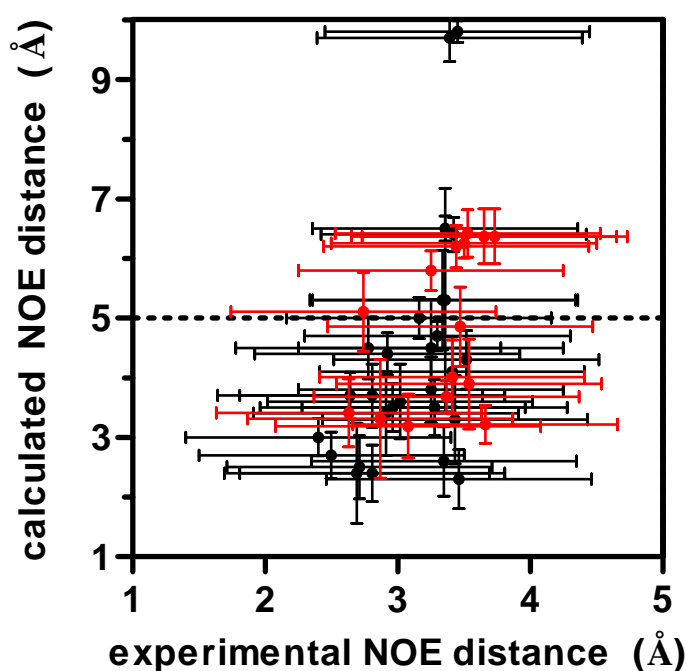
Figure 3.5: Correlation of experimental and back-calculated distances from NOE. Red symbols, restraints from rigid residues; black symbols, restraints from flexible residues. Horizontal error bars represent ±1 Å. Vertical error bars represent 1× standard deviation for the calculated structures. Horizontal dashed line represents a distance of 5 Å.

3.4   Conclusions

We have successfully determined the binding pose of ligand **1** bound to FKBP12 by intermolecular NOE restraints. Not all NOEs were satisfied in the final structure, indicating a certain extent of dynamics. Nevertheless, the final structure is convergent among different PDB structures and all the NOE crosspeaks indicate that ligand **1** binds at site 1. This final structure is used in Chapter 4 as a reference for an orthogonal comparison for the PCS-derived structure.

4

Structure determination of a protein-ligand complex by paramagnetic NMR spectroscopy

Abstract

Determining the three dimensional structure of a small molecule-protein complex with weak affinity can be a significant challenge. We present a paramagnetic NMR method to determine intermolecular structure restraints based on pseudocontact shifts (PCSs). Since the ligand must be in fast exchange between free and bound state and the fraction bound can be as low as a few percent, the method is ideal for ligands with high micromolar to millimolar dissociation constants. Paramagnetic tags are attached, one at a time, in a well-defined way *via* two arms at several sites on the protein surface. The ligand PCSs were measured from simple 1D ^{1}H spectra and used as docking restraints. An independent confirmation of the complex structure was carried out using intermolecular NOEs (Chapter 3). The results show that structures derived from these two approaches are similar. The best results are obtained if the magnetic susceptibility tensors of the tags are known, but it is demonstrated that with two-armed probes, magnetic susceptibility tensor can be predicted with sufficient accuracy to provide a low-resolution model of the ligand orientation and the location of the binding site in the absence of isotope labeled protein. This approach can facilitate fragment-based drug discovery in obtaining structural information on the initial fragment hits.

4.1 Introduction

The availability of three dimensional (3D) structures of protein-ligand complexes significantly improves the efficiency of refining hits towards leads in the early stages of drug discovery.[174] Typically, structure-driven hit optimization programs rely on crystallographic data. In fragment-based drug discovery (FBDD), however, weakly binding ligands (fragments) are often not observed in crystals, for a variety of reasons. NMR spectroscopy is a powerful alternative for deriving structural information, particular for weakly interacting complexes. The classical NMR approach is based on the observation of intermolecular NOEs and is suitable for most proteins smaller than 40 kDa. Although robust, the method can be time consuming and requires uniformly isotopically labeled protein. The method is strictly limited to proteins that can be functionally expressed under these conditions. Selective isotope labeling schemes have been employed in combination with deuteration to enable NMR analysis of much larger proteins. Complexation-induced chemical shift perturbation (CSP) data have been used as ambiguous interaction constraints to calculate structures of protein–ligand complexes.[175] However, this method monitors both direct and remote effects, and therefore the .binding site is not always well defined. We sought a method that could in principle be applied to proteins where no, or only limited, isotope labeling can be performed.

Paramagnetic NMR is known to be a powerful tool to study biological systems due to its versatile effects, including pseudocontact shifts (PCSs), paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC).[89–91] Paramagnetic NMR has been applied extensively to characterize protein-protein interactions, but very little in protein-small molecule interactions. Pioneer studies used the PRE caused by a paramagnetic metal ion[95] or spin label[92,93] to facilitate ligand screening and potentially, to obtain information on the ligand pose.[95]

A combination of PCS and RDC was applied to assist structure determination of a carbohydrate-protein complex where the paramagnetic center was introduced by creating a fusion protein with a C-terminal lanthanide-binding peptide tag (LBT).[96] PRE assisted ligand docking with a spin-labeled peptide bound specifically to a protein was also reported.[98] Recently, paramagnetic effects stemming from a two-point anchored N-terminal LBT were used to determine the structure of protein-peptide complexes.[99] PCSs have also been applied to determine the structure of a small molecule ligand in rapid exchange with a protein in which a lanthanide was bound in a natural metal binding site.[97] In this study the ligand bound very closely to the lanthanide, enabling the use of large ligand-to-protein ratios. This

latter example is the only case where PCSs have been used to elucidate the structure of a protein-ligand complexes containing a small fragment (M_w <300 Da).

We demonstrate here an alternative way to determine the location and orientation of a weakly bound fragment with PCS restraints from a ligand in rapid exchange with a protein. A rigid, double-armed lanthanide-binding tag, CLaNP-5,[144,176] was attached at three different sites, one at a time, on the protein surface *via* disulfide bond linkage. Using this tag, the magnitude and orientation of the Δχ-tensors can be predicted. Further, the paramagnetic effects can be tuned by using different lanthanides in the tag. We selected a 12 kDa FK-506 binding protein (FKBP12) as a model protein to investigate the potential of the methodology. FKBP12 is a peptidyl-prolyl isomerase which belongs to the family of immunophilins and is a drug target for the immuno-suppressants rapamycin and FK506. There have been many structural studies on this protein.[9,10,15,20,162,163] The ligand in this study is a fragment that was identified as a hit against FKBP12 from a screen of a fragment library using Target-Immobilized NMR Screening (TINS).[84] In this work, we compare the PCS-based docking result with the structure determined by intermolecular NOEs (Chapter 3). The structure calculations were performed using parameters from both predicted and experimentally determined paramagnetic Δχ-tensors. The result shows that, even without resonance assignments of the protein, it is possible to determine the ligand binding site and approximate orientation. The method can assist lead optimization in fragment-based drug discovery when high resolution structural information is not available.

4.2   Materials and Methods

Ligand preparation

Assignments and preparation of the ligand **1** (Figure 3.1A) was carried out as described in Chapter 3.

Protein expression and purification

Uniformly ^{15}N-labeled FKBP12 double cysteine mutants (K34C/K35C, K44C/K47C, C22V/E61C/Q65C), all of which contain an additional LEHHHHHH tag at the C-terminus were purified from *Escherichia coli* strain BL21(DE3) containing the overexpression plasmid pET20b with FKBP12 insert. The expression and purification are described in Chapters 2 and 3. The yield was in general 15-23 mgL[-1] for the double cysteine mutants.

CLaNP-5 attachment

The details of CLaNP-5 attachment to FKBP12 variants are described in Chapter 2.

NMR measurements

All protein NMR samples contained 15 mM Tris-HCl, 25 mM NaCl, pH 7.7, 6% D_2O for ^{15}N-labeled protein and >95% D_2O for non-isotope-labeled protein. The concentration of Ln^{3+}-CLaNP-5 attached FKBP12 was 17 to 39 μM. The ligand to protein ratios were 1.3:1 for all Ln^{3+}-CLaNP-5 attached FKBP12 mutants. 1D-^1H and [^1H,^{15}N]-HSQC spectra were recorded at 290 K on a Bruker Avance DMX-600 spectrometer equipped with a TCI-Z-GRAD cryoprobe. 1D-^1H spectra of the complex CLaNP5-FKBP12 and ligand **1** were recorded with a spectral width of 16 ppm and 6 k complex points, resulting in a digital resolution of 1.56 Hz before zero filling. Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence was used with a total echo time of 60 ms comprising 60 pulses, for suppression of macromolecules resonances. Data were processed in TopSpin (Bruker) and then spectra were analyzed in Sparky.[164]

Magnetic susceptibility tensor optimization and PCS-based structure calculations

PCSs are defined as the difference in ppm between the corresponding resonance in the paramagnetic sample and the diamagnetic sample. The PCS gives information on the distance and angle between a nucleus and the paramagnetic center according to Equation 1.3.[113] Errors were calculated by randomly excluding 10% of the data with Monte Carlo analysis implemented in Numbat.[177]

The fits of observed versus back-calculated protein PCSs and docking were performed in the XPLOR-NIH[178] program containing the PARArestraints module.[179] The predictions of the initial Δχ-tensor positions and orientations were carried out as previously described.[144] The structure model, parameter and topology files of the ligand were generated from PRODRG server.[153] Bound ligand PCS values were used as the experimental restraints for PCS docking. PCSs of ligand methyl protons were used for methyl carbon positions instead of methyl proton positions. The protein backbone atoms and the pseudo-residues defining the metal coordinates were fixed in the docking process. The three datasets of the different tag positions on FKBP12 were used simultaneously. Ligand PCSs from all three tagging sites were used in the docking procedure with equal weighting. Each docking calculation comprised 100 steps, started with restrained rigid body docking for 20 steps (0.01 pico-seconds increments, 300 increment evaluations/step). The lowest energy structure was

subsequently subjected to restrained Langevin dynamics for 80 steps (0.001 pico-seconds increments, 2000 increment evaluations/step), which allowed the ligand and the residues within 8 Å from the ligand to be flexible. 200 independent docking calculations were performed using the solution structure of FKBP12 (PDB entry 1FKR, model 14)[9] and the predicted $\Delta\chi$-tensor parameters and position of the lanthanide ion. The same calculations with experimentally determined $\Delta\chi$-tensor parameters and positions were performed for comparison. The agreement between the experimental PCSs and back-calculated PCSs was evaluated using the Q factor defined in Equation 4.1 :[180]

$$Q = \sqrt{\frac{\sum_i \left(PCS_i^{\text{exp}} - PCS_i^{calc}\right)^2}{\sum_i \left(\left|PCS_i^{\text{exp}}\right| + \left|PCS_i^{calc}\right|\right)^2}} \qquad (4.1)$$

where PCS_i^{exp} and PCS_i^{calc} are the observed and calculated PCSs. The XPLOR-NIH script is available in Appendix B.

PCSdock prediction

The PDB structure containing all three $\Delta\chi$-tensors (calculated or predicted) was imported into PCSdock using Scientific Python.[181] Ligand PCS data were imported, along with the axial and rhombic magnitudes of the $\Delta\chi$-tensors. A cubic grid of user-defined size (30 Å was used) and 1 Å spacing was placed around the protein using the centre of mass as its origin. Predicted PCS values for the points on the grid were calculated for each paramagnetic $\Delta\chi$-tensor. The experimental ligand PCS values for an atom *j* were then compared to the predicted PCS values for each grid point *i* using a Q score defined by Equation 4.2:

$$Q_{i,j} = \frac{1}{N} \sum_{k=1}^{N} \sqrt{\frac{(PCS_{i,k}^{pred} - PCS_{j,k}^{exp})^2}{(\left|PCS_{i,k}^{pred}\right| + \left|PCS_{,j,k}^{exp}\right|)^2}} \qquad (4.2)$$

where *N* is the number of paramagnetic $\Delta\chi$-tensors, $PCS_{i,k}^{pred}$ is the PCS calculated for the grid point *i* relative to the paramagnetic $\Delta\chi$-tensors *k*, $PCS_{j,k}^{exp}$ is the experimental PCS used for the atom *j* relative to the paramagnetic $\Delta\chi$-tensors *k*. If the $Q_{i,j}$ for any atom of the ligand was lower than a user-defined level, the grid point *i* was accepted. If the $Q_{i,j}$ was larger for all atoms of the ligand, that grid point was discarded. The PCSdock script is kindly provided by Simon P. Skinner (Leiden University).

4.3 Results

Selection of CLaNP-5 tagging sites

In order to determine the protein-ligand complex, we also utilized paramagnetic NMR and generated three FKBP12 mutant proteins tagged with the synthetic lanthanide tag CLaNP-5. The lanthanide tag, CLaNP-5, is designed to covalently link to the protein surface *via* two disulfide bridges.[144,176] Therefore, the presence of surface accessible cysteine pairs is essential. Several criteria need to be considered for selecting the mutation sites: (1) The cysteines should be far enough from the putative ligand binding site to avoid interference with binding, yet close enough to yield appreciable paramagnetic effects; 15-30 Å is a reasonable estimation based on the location of the binding site and the total size of FKBP12. 25 Å is an estimation of the effective range for Yb^{3+}-chelated CLaNP-5. In this study, the distances between the putative ligand binding site and the three lanthanide tags are in the range of 15-25 Å. It is possible to use other lanthanides to adjust the effective range for different protein sizes.[117] (2) The two $C\alpha$ atoms from the cysteines should be 6-10 Å apart, with their side chains pointing away from the protein surface and roughly in the same direction. Cysteines buried inside the protein cannot react with CLaNP-5. The locations of cysteine mutation sites (K34C/K35C, K44C/K47C and E61C/Q65C) were selected to satisfy the above requirements. As a consequence of the anisotropy of the magnetic susceptibility, the ligand PCSs can be close to zero regardless of the distance if the ligand is located close to the region in which the PCS changes its sign. Nevertheless, measuring PCSs in this situation for a different tag position will provide information. This is one of the advantages of synthetic tags compared with the methods using lanthanide-binding peptides[182,183] or metal displacement, in which usually only one tag position is available.

Ligand PCS measurement

The resonance assignments of the free ligand **1** were obtained by analysis of 1D-^{1}H, ^{13}C-APT, [^{1}H,^{13}C]-HSQC and [^{1}H,^{13}C]-HMBC spectra (Chapter 3). The binding affinity of the ligand **1** to FKBP12 was determined using [^{1}H,^{15}N]-HSQC titrations, and was found to be 1.80 (±0.05) mM. (Chapter 3). To eliminate interference from the protein resonances, 1D-^{1}H NMR spectra of the ligand in the presence of Ln^{3+}-CLaNP-5-FKBP12 were recorded using a T_2 relaxation delay of 60 ms. Ligand PCSs were measured from singly tagged protein with Yb^{3+}-CLaNP-5 in the paramagnetic sample and Lu^{3+}-CLaNP-5 in the diamagnetic sample. In total, 3 different pairs of protein samples, each containing the CLaNP-5 attached at a

different site, were used. Initially **1** was titrated into Yb^{3+}-CLaNP-5-FKBP12 to determine the optimal molar ratio of protein to ligand. The ligand and protein concentrations were 36 and 28 µM (34C/35C), 24 and 17 µM (44C/47C), and 47 and 39 µM (61C/65C). Under these conditions 0.9-1.9 % of the ligand is bound. A protein to ligand ratio 1:1.3 was used for all measurements with both Yb^{3+}-CLaNP-5 and Lu^{3+}-CLaNP-5, because this ratio represented the best compromise between size of the PCSs, the line broadening and spectral resolution. Figure 4.1 shows the spectra from which the ligand PCSs were determined for the 34C/35C tagging site.
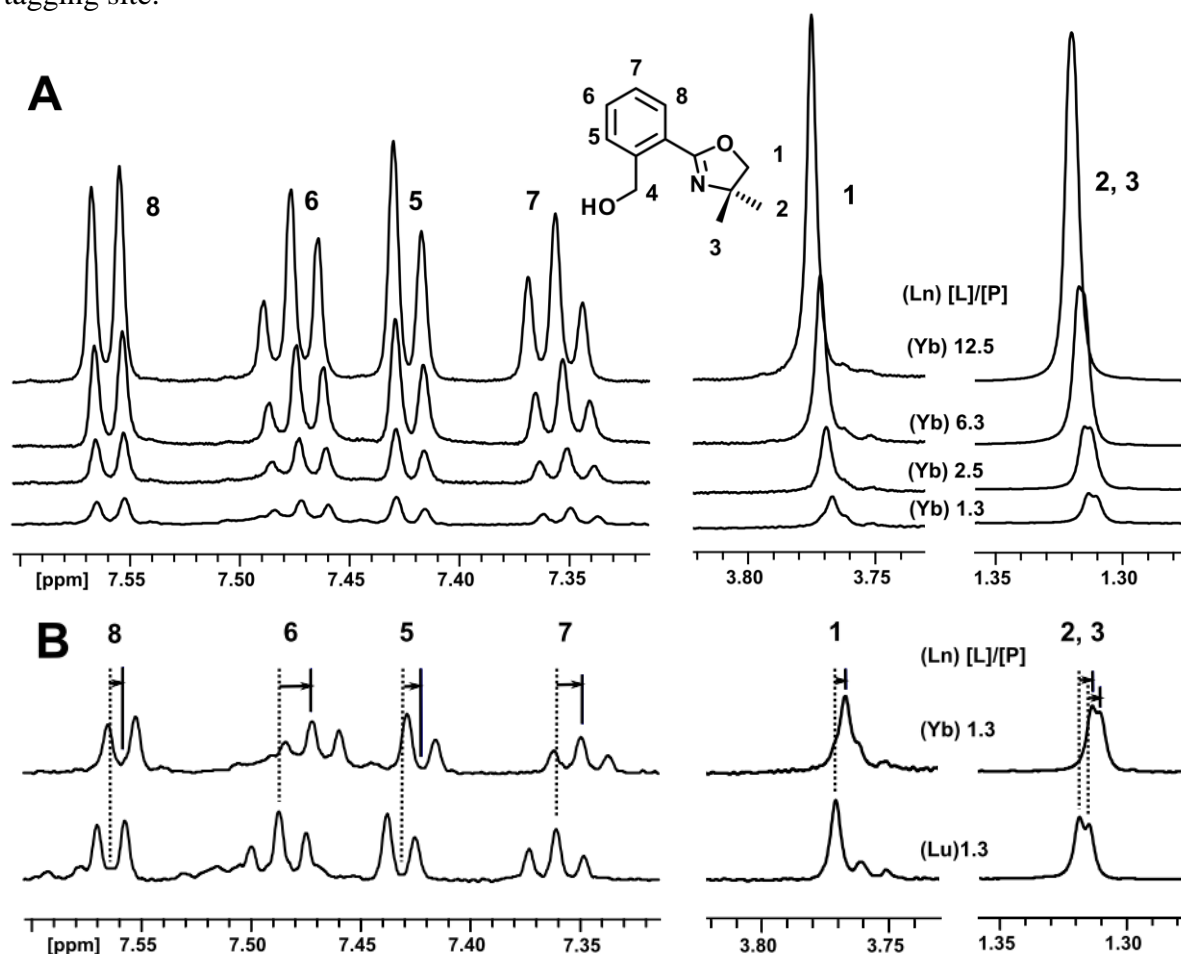
Figure 4.1: (A) Overlay of 1D-^{1}H NMR spectra of ligand **1** in the presence of 28 µM FKBP12 (34C/35C) attached to paramagnetic Yb^{3+}-CLaNP-5 with increasing ligand/protein molar ratios, which are indicated in the spectra. Proton assignments of **1** are indicated by corresponding numbers on the structure. The chemical shift of proton 4 overlapped with the water resonance and is not shown here. The resonances of the methyl groups numbered 2 and 3 are degenerate in the free form but are resolvable in the bound form. (B) The observed ligand PCS is the difference between the resonance positions for the paramagnetic (Yb^{3+}) and diamagnetic (Lu^{3+}) samples. The dashed and solid lines indicate the positions of the diamagnetic and paramagnetic ligand resonances, respectively. Spectra were recorded at 600 MHz.

The observed ligand PCS values are small (less than 10 Hz) due to the small fraction of bound ligand (1.5%). However, they can be measured precisely because of the sharp signals of the ligand. The PCSs are summarized in Figure 4.2. A control measurement of the ligand in the presence of the free probe in the absence of protein showed no paramagnetic effects (data not shown). This indicates that the observed ligand PCSs derive exclusively from the bound state of the ligand. A total of 21 PCS values were obtained from the three pairs of para- and diamagnetic spectra.
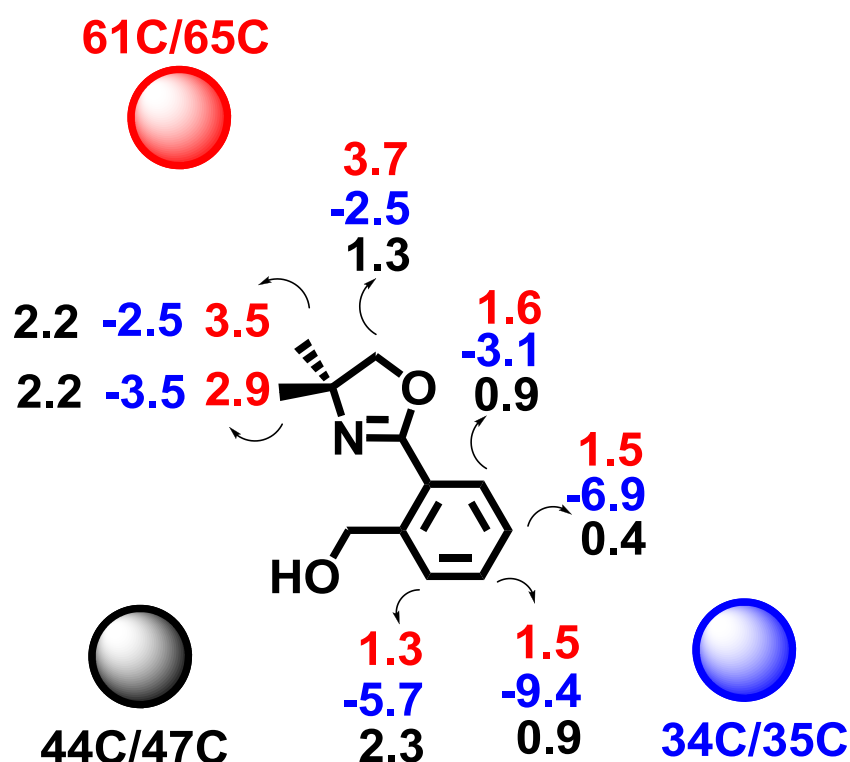
Figure 4.2: Observed ^{1}H PCS values (in Hz, for spectra acquired at 600 MHz) of ligand **1** from three different locations of the paramagnetic center. The relative positions of the paramagnetic centers are indicated and the associated PCS values are color coded. Blue, PCSs from 34C/35C; black, PCSs from 44C/47C; red, PCSs from 61C/65C.

PCS-based structure calculation with predicted $\Delta\chi$-tensors

For ligands in fast-exchange on the NMR time-scale (i.e. $\Delta\Omega < k_{off}$), the observed PCSs are weighted averages of the free and bound states. Therefore, the PCS values for the bound state can be derived if the fraction of ligand bound is known. Using the known concentration of ligand and protein, as well as the experimentally determined K_D, the observed PCSs were converted to the PCSs in the bound state (Table 4.1).

Table 4.1: Ligand pseudocontact shifts converted to 100% bound (ppm) according to calculated bound fractions (indicated at bottom of the table). The observed values are given in Figure 4.2. Uncertainties (parentheses) are estimated as $\pm 2\times$ standard deviation of the positions of the para- and diamagnetic resonances based on the following formulae: $\sigma = \sqrt{\sigma_{para}^2 + \sigma_{dia}^2}$

Spin	K34C/K35C	K44C/K47C	C22V/E61C/Q65C
H-1	-0.28 (0.06)	0.23 (0.08)	0.32 (0.04)
H-2	-0.29 (0.02)	0.40 (0.02)	0.31 (0.06)
H-3	-0.33 (0.02)	0.40 (0.02)	0.26 (0.04)
H-5	-0.63 (0.12)	0.42 (0.18)	0.11 (0.10)
H-6	-1.04 (0.38)	0.16 (0.34)	0.13 (0.18)
H-7	-0.77 (0.24)	0.07 (0.32)	0.14 (0.18)
H-8	-0.34 (0.12)	0.17 (0.13)	0.14 (0.01)
[Protein] (μM)	28	17	39
[Ligand] (μM)	36	24	47
Bound fraction (%)	1.5	0.9	1.9

Previous work in which CLaNP-5 was bound to a rigid protein suggested that lanthanide position and the $\Delta\chi$-tensor orientation can be predicted using a simple set of rules.[144] This characteristic is important for proteins that cannot be isotopically labeled, so we first determined the structure of the complex of **1** bound to FKBP12 using the predicted $\Delta\chi$-tensor orientations. As with the NOE-based structure determination, several structures of FKBP12 from the PDB were used as input. PDB file 1FKR model 14 is shown as an example. In the docking procedure, the ligand was allowed to move, and the protein was fixed, except for side chains within 8 Å of the ligand, which were allowed to rotate. The complex formed by FKBP12 and ligand **1** was energy minimized guided by the energy terms for the pseudocontact shifts and Lennard-Jones potential. The five lowest energy structures were selected, based on the total energy and overlaid with the structure derived from NOE restraints (Figure 4.3). Two orientations are found with low PCS energy, which differ by a rotation of ~90 degrees (Figure 4.3, A and B), and the lowest energy cluster (Figure 4.3A) is closest to the orientation observed in the NOE structure. The average RMSD relative to the NOE structure is 4.7 ± 0.9 Å. In the other orientation, which has on average 10% higher PCS energy, the aromatic ring of **1** points outwards from the binding site (RMSD relative to NOE 4.4 ± 0.1 Å).
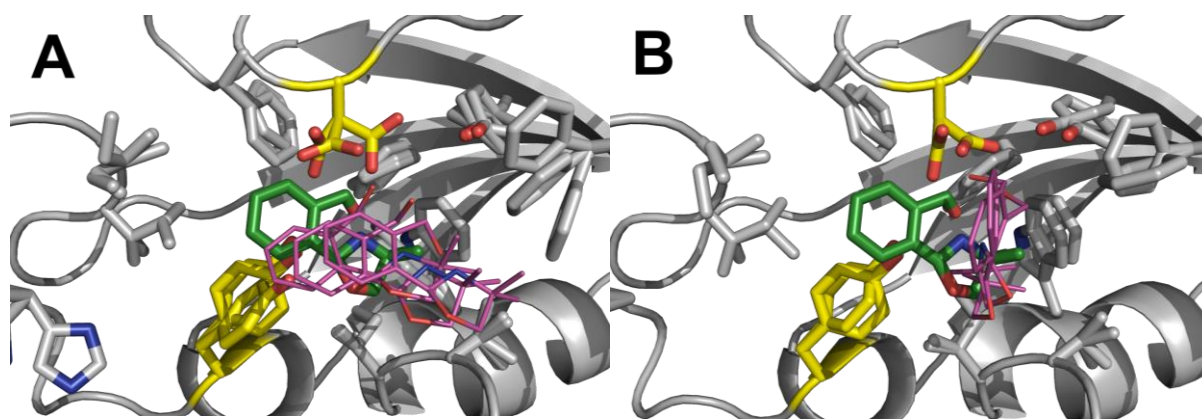
Figure 4.3: Best five PCS structures calculated using the predicted Δχ-tensors (ligands in magenta) superimposed on the averaged NOE structure (ligands in green). Two clusters with similar PCS energy are present with the ligand rotated by ~90 degrees. (A) The lowest energy cluster has an orientation parallel to the NOE structure. (B) The second cluster has the aromatic ring pointing out from the binding site.

By applying the predicted Δχ-tensor parameters for PCS-based structure calculations, the approximate location of the ligand binding site could be established. Both clusters of calculated orientations showed good agreement between the predicted and experimental ligand PCSs (Figure 4.4, A and B), as indicated by the quality factor (equation 4.1). This suggests that, using CLaNP-5 with predicted Δχ-tensor parameters, it is possible to detect the ligand binding site and obtain a low resolution structure for further ligand optimization. For this approach, prediction of the Δχ-tensor position and orientation does not require any PCSs or other NMR data on the protein, although the 3D structure of the protein is required.

Δχ-Tensor calculations

PCS values depend on the position of the nuclear spin relative to the magnetic susceptibility anisotropy tensor of the lanthanide ion, as described by equation 1.3. The Δχ-tensors are defined by eight parameters (the axial and rhombic components, the x, y, z coordinates of the metal, and the orientation of the Δχ-tensor relative to the molecular frame, defined as three Euler angles). Therefore, they can be determined from a minimum of eight PCSs measured from the nuclear spins with known resonance assignment. The PCS values of protein amide protons were measured as the difference in the chemical shift between the Yb^{3+}- and Lu^{3+}-CLaNP-5 attached FKBP12. Most of the [^{1}H,^{15}N]-HSQC assignments of the wild type FKBP12 could be transferred to the double cysteine mutants attached to the Lu^{3+}-CLaNP-5.
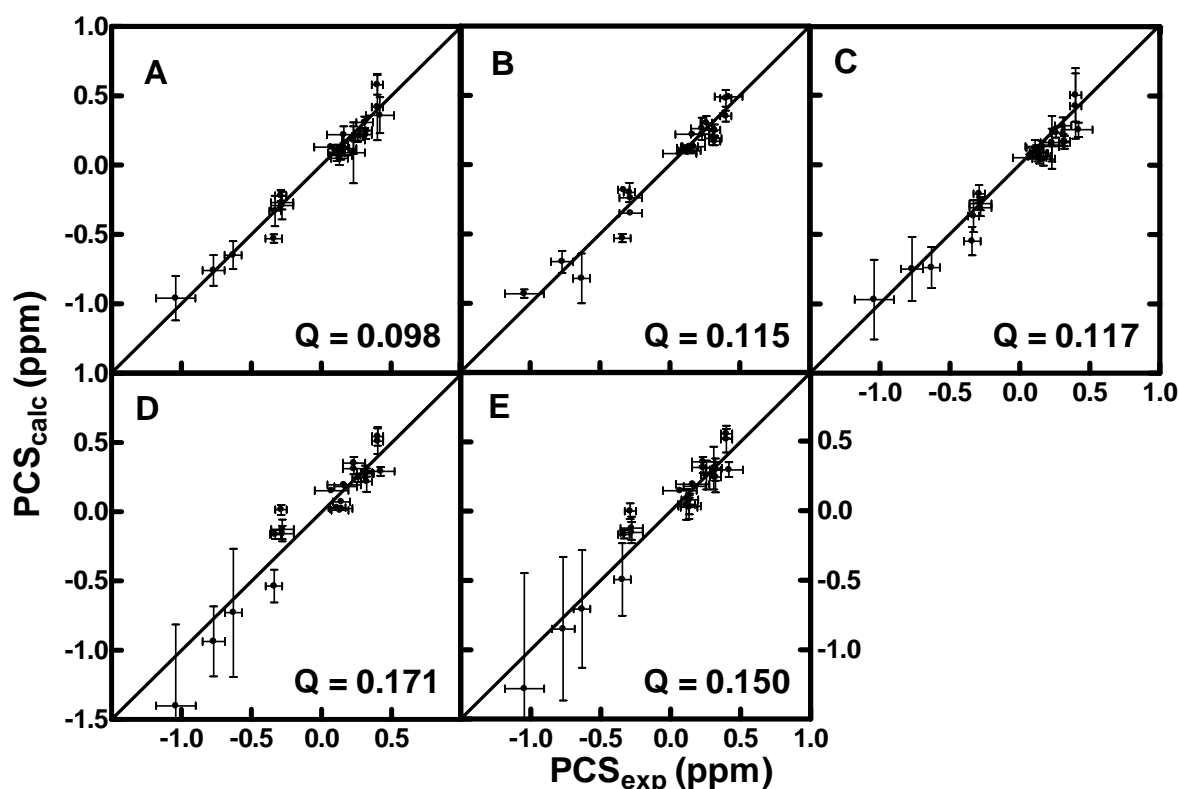
Figure 4.4: Correlation between experimental and back-calculated ligand PCSs for the top five structures of **1** bound to FKBP12 as determined from predicted and experimentally determined $\Delta\chi$-tensor positions. (A) Correlation of the best structures obtained with predicted $\Delta\chi$-tensors, lowest energy cluster, as shown in Figure 5A and (B) second lowest energy cluster, as shown in Figure 5B. (C) Best five structures with experimentally determined $\Delta\chi$-tensors. (D) Back-calculated ligand PCS values for NOE-based structures using predicted $\Delta\chi$-tensor positions. (E) Back-calculated ligand PCS values for NOE-based structures using experimentally determined $\Delta\chi$-tensor positions. Vertical error bars represent 2× standard deviation of the variation in the sets of structures for the back-calculated PCSs. Horizontal error bars represent the estimated experimental error.

Although there are ambiguous assignments in the spectra of the mutants, more than 70% of the residues could readily be assigned and therefore the $\Delta\chi$-tensor magnitude and orientation could be determined for each of the three mutants. The calculated parameters are summarized in Table 4.2. Overlays of the paramagnetic and diamagnetic [^{1}H,^{15}N]-HSQC spectra of the three mutants and the correlations between the experimental and back-calculated protein PCSs are shown in Figure 4.5. Compared with FKBP12 variants K34C/K35C and K44C/K47C, C22V/E61C/Q65C exhibits somewhat smaller tensor magnitudes (Table 4.2) and a slightly poorer fit (Figure 4.5), which is an indication that the tag at this position is a little more mobile than at the other two positions. Nevertheless, all

the values of the axial and rhombic components from three tagging sites are comparable to previously reported values,[144,184–188] and therefore the values for C22V/E61C/Q65C were considered reasonable.

Table 4.2: $\Delta\chi$-Tensor parameters of Yb^{3+}-CLaNP-5 attached to FKBP12.[a]

Mutant	K34C/K35C	K44C/K47C	C22V/E61C/Q65C
$\Delta\chi_{ax}$ ($10^{-32}m^3$)	8.7 ± 0.4	8.9 ± 0.2	7.8 ± 0.1
$\Delta\chi_{rh}$ ($10^{-32}m^3$)	3.4 ± 0.4	3.3 ± 0.3	2.7 ± 0.2
Q (eqn. 4.1)	0.023	0.021	0.035
Restraints	75	73	72

[a] Errors were estimated by randomly excluding 10% of the data with a Monte Carlo approach.[177]

PCS-based structure calculation with experimentally determined $\Delta\chi$-tensors

The docking of ligand **1** to FKBP12 was repeated using the experimentally determined tensor orientations. The orientations of the best five structures form a single cluster that is close to the NOE-derived structure (Figure 4.6). The correlation of experimental and back-calculated ligand PCSs is presented in Figure 4.4C. The average RMSD between the cluster members is 1.6 ± 0.5 Å and to the averaged NOE structure 2.8 ± 0.4 Å. With the experimental $\Delta\chi$-tensor parameters, the quality factor has slightly increased to 0.117. Given the error estimates shown in Figure 4.4, we conclude that the calculated structures fit the data and that the differences of the Q values between the clusters derived from the predicted and experimental $\Delta\chi$-tensors are not significant. The position of the ligand found with the optimized $\Delta\chi$-tensors is closer to the NOE-based position, suggesting that the optimization increases the accuracy of the solution. The PCSs were also back-calculated from the NOE-based structures, using both the predicted $\Delta\chi$-tensors (Figure 4.4D) and the experimentally determined $\Delta\chi$-tensors (Figure 4.4E). It is clear that the NOE-derived ligand position fits the PCS data worse than the calculated position, indicating that the two positions differ significantly from the experimental point of view. The NOE-based structure is used here as the standard for validation of the PCS based structure, but, as mentioned in Chapter 3, the NOE-based solution is also an approximation due to the dynamics in the binding site.
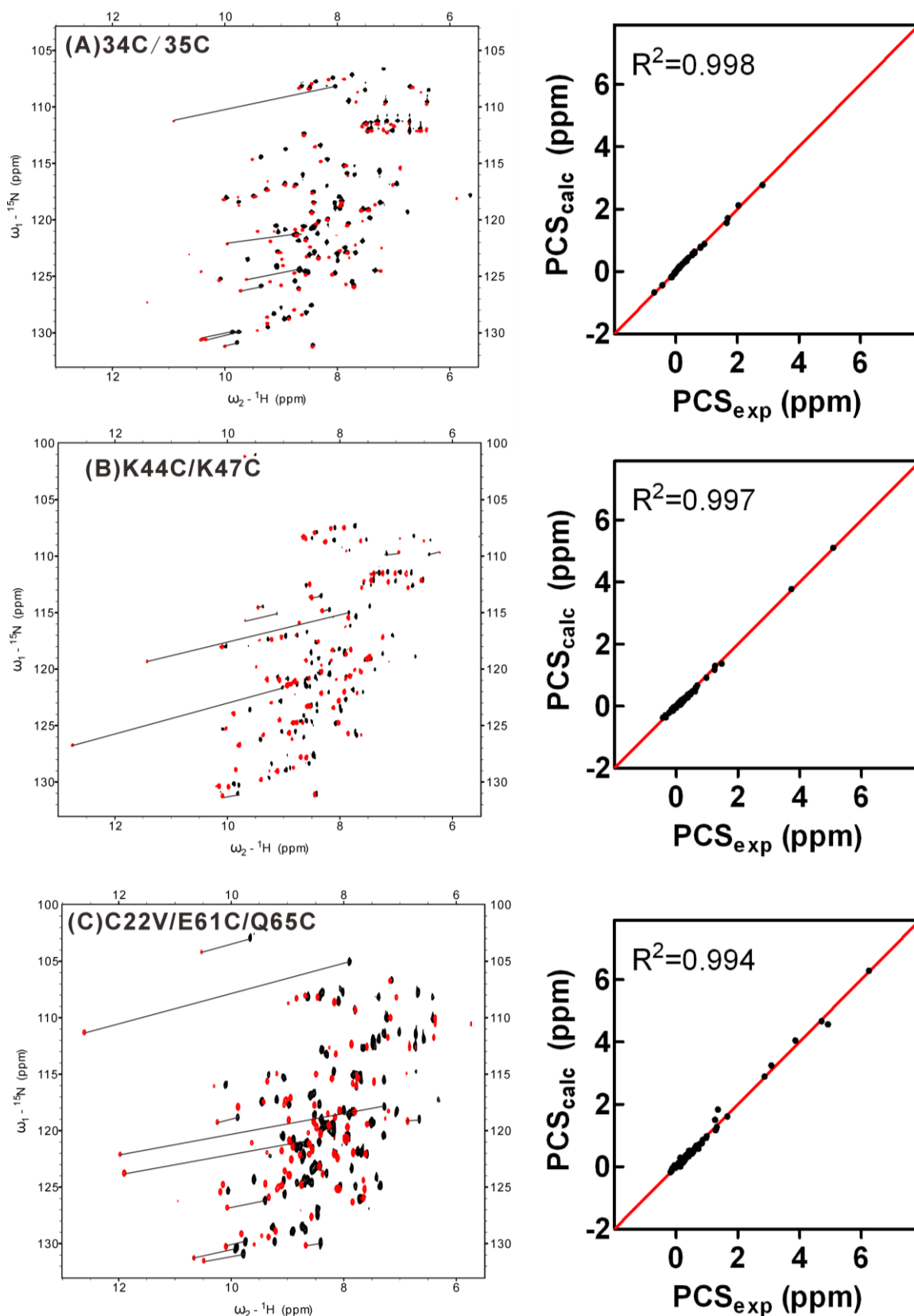
Figure 4.5: (Left) [^{1}H, ^{15}N]-HSQC overlay of CLaNP5 attached FKBP-12 mutants with Yb (red) and Lu (black). Solid lines show examples of PCS from selected corresponding residues. (Right) Correlation between experimental PCS and back-calculated PCS from FKBP-12 mutants attached to CLaNP-5(Yb). (A) K34C/K35C, (B) K44C/K47C, (C) C22V/E61C/Q65C.
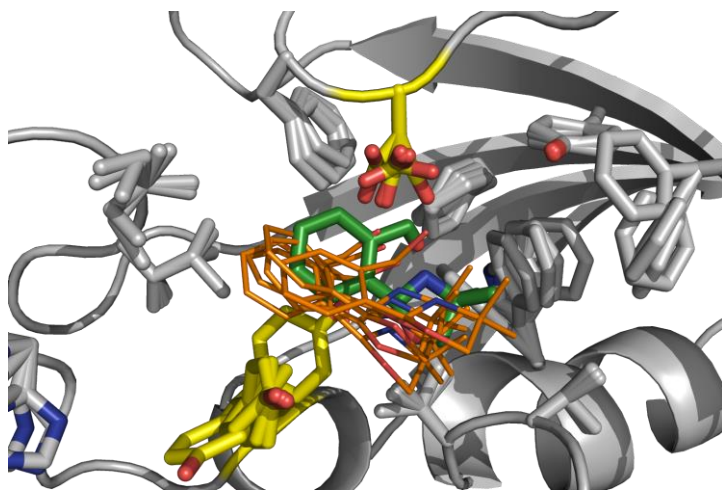
Figure 4.6: Superposition of the averaged NOE structure (in green) of the FKBP12-1 complex and the best five structures (in orange) calculated using ligand PCSs and the three experimentally determined $\Delta\chi$-tensors. The protein backbone is represented as a grey ribbon except for the residues D37 and Y82 (in yellow). The average RMSD of the ligand from PCS calculations relative to the NOE calculation is 2.8 ± 0.4 Å.

A 'ghost' site found by PCS due to degeneracy of $\Delta\chi$-tensors

A commonly encountered problem in paramagnetic NMR is the observation of multiple structure solutions due to the degeneracy of $\Delta\chi$-tensor frames.[189-191] Here, we encountered a similar situation. In the calculations, another cluster of ligands was found at a site entirely different from the one shown in Figure 4.6 and which was neither identified by any intermolecular NOE, nor reported in the literature. Using the predicted $\Delta\chi$-tensor parameters, this position was not identified. The position is located near a loop consisting of amino acids S39 to F46, opposite to the binding site mentioned above (Figure 4.7). At this 'ghost' site, the ligand has no van der Waal's contacts with the protein and is exposed on the protein surface and therefore it does not appear to be physically realistic. This non-physical position is apparently an artifact which originates from the degenerate nature of the $\Delta\chi$-tensors. By removing solutions that have zero van der Waal's energy, it is possible to eliminate this artifact.
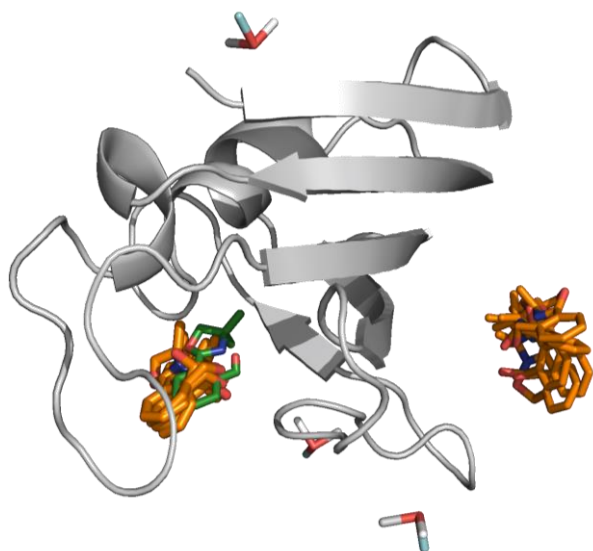
Figure 4.7: Second ligand position is defined by experimentally determined tensors. The two clusters calculated using the experimentally determined tensors are shown in orange and the averaged NOE position is shown in green. The metal positions and tensor orientations are shown in sticks, with the z-axis in cyan.

In order to visualize the degeneracy in a set of three tags, a script named PCSdock was written to produce a grid around the protein and calculate the PCS values of the pre-defined tags for each grid point. The algorithm then compares the experimental PCSs with the three calculated ones and calculates a Q value (equation 4.2). If the Q value for any atom of the ligand is below a threshold, that grid point is selected, otherwise it is discarded. Thus, PCSdock gives a fast, though crude, representation of where the ligand could be located on the basis of the PCSs. While the procedure is likely to overestimate the number of possible locations, it can be a useful tool for establishing degeneracy. Applied to FKBP12 with the three tags and the experimentally observed PCSs, the calculation clearly shows two possible, spatially distinct areas (Figure 4.8A). Each of the two low energy clusters from the calculation using experimentally determined Δχ-tensors fits into one of the areas defined by the PCSdock calculation. Interestingly, this 'ghost' site was not found when calculated using the predicted Δχ-tensors (Figure 4.8B).
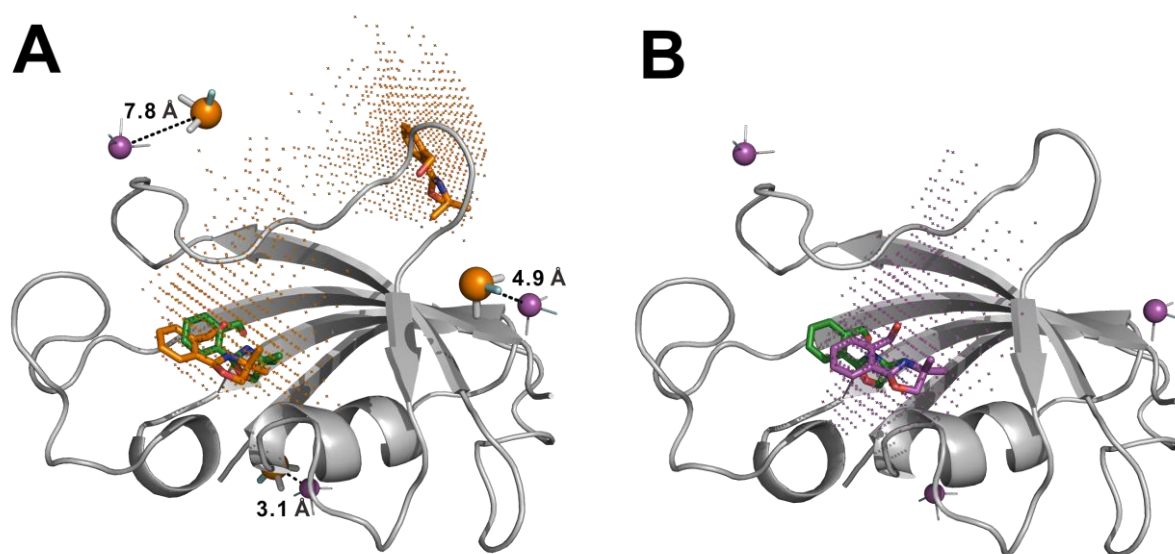
Figure 4.8: Grid points produced by PCSdock (Q = 0.15). The metal positions are shown in spheres (magenta, predicted positions; orange, experimentally determined positions), and Δχ-tensor orientations are shown in sticks, with the z-axis in cyan. Ligands in green are the NOE-based position. (A) The two ligands in orange show the two locations determined by the experimental Δχ-tensors. Two clusters were determined by PCSdock (orange dots): one close to the actual binding site, the other at the 'ghost' position. Distances between the metal positions determined by pure prediction and by experimental PCS are indicated. (B) Ligand in magenta shows the position determined using the predicted tensors. Only one cluster was found by PCSdock (magenta dots).

4.4   Discussion

We have demonstrated a general method to obtain structural information on complexes of weakly binding, small molecule ligands and proteins. The method seems most applicable to proteins that cannot be isotopically labeled, for which X-ray crystallography fails and for which many ligands need characterization. Although powerful, the method has certain limitations. On the target side, proteins with many surface-exposed cysteine residues are not appropriate as it would be quite challenging to attach the CLaNP-5 tag at one unique position. There are also restrictions on the ligands such that those with symmetric structures, with many scalar couplings or that exhibit intermediate exchange on the NMR time scale can cause difficulties in analyzing the data. Ligands that have symmetric structures tend to show degeneracy in NMR spectra, resulting in a reduced number of restraints that are themselves degenerate. Ligands with many scalar couplings, such as a saturated hydrocarbon ring, have ^{1}H spectra that can be difficult to resolve and the PCSs may be difficult or impossible to measure. Similarly, ligands in intermediate-exchange are difficult to study in NMR due to NMR line broadening leading to resonance overlap or less visible resonances. It should be noted that such ligands will present a challenge to any paramagnetic NMR approach, not just those based on the use of a lanthanide tag.

If the goal is to precisely define the binding site of the ligand on the protein, it is essential to obtain a sufficient number of meaningful restraints. Due to their small volume, fragments can only generate restraints covering a limited part of $\Delta\chi$-tensor space, easily leading to degenerate solutions. We addressed this issue by attaching CLaNP-5 at several sites throughout the protein surface. Three tag positions were used to generate a total of 21 PCS restraints for structure calculations. It is also possible to use CLaNP-5 with different lanthanides at a single tagging site, an approach used in previous studies for which multiple $\Delta\chi$-tensor positions were not available.[97,99] However, the tensor orientations of different lanthanides are similar in the same tag, so the PCSs merely scale with the magnitude of the anisotropy, reducing the information content of the extra restraints. In this case, a combined application of CLaNP-5 and CLaNP-7[188] would be possible using the same tag position, because the $\Delta\chi$-tensor orientations differ between these two CLaNP molecules. We expected that the three tagging sites would lead to a unique solution for the binding site. Unfortunately, degeneracy remained, leading to a physically unrealistic binding site in addition to the correct site. While the relevant site could be readily distinguished from the irrelevant one, some caution must nevertheless be exercised when interpreting the results. The presence of

this 'ghost' site suggests that, depending on the actual situation, more than three mutants might be required to fully break all degeneracy. Additionally, it is important to consider the (potential) dynamic behavior of sites within the protein when selecting positions to generate the dual cysteine mutations. Motion of the attachment site within the protein can influence both the position of the paramagnetic center and magnitude of the $\Delta\chi$-tensor.[176,185] For example, the large difference between the predicted and experimental position of the lanthanide for FKBP12 (K34C/K35C) and the orientation of its tensor (Figure 4.4A and Figure 4.5) is likely related to its location on an ill-defined protein loop, which makes the prediction less reliable and bears the risk that the tag slightly affects the average structure of the loop.

A number of NMR approaches have been proposed to determine the structure of protein-small molecule complexes. The approach based on intermolecular NOEs, which is robust and provides structural detail, is most frequently used. The method emphasizes short-range restraints (<5 Å or <8 Å depending on the isotopic labeling scheme) and requires resonance assignments of the protein. With the assistance of computational modeling, the NOE method can be relatively fast, as demonstrated in the system of matrix metalloproteinases (MMPs) and its tight-binding ligands.[192] A particular limitation of the method is the requirement for isotope-labeled protein, typically necessitating expression in *E. coli*. In contrast, methods that exclusively observe the NMR spectrum of the ligand can, in principle, be applied to protein derived from nearly any source. Epitope mapping, a method that quantitates the amplitude of ligand resonances in a saturation transfer difference spectrum (STD),[76,77] can provide information on how the small molecule binds to the protein. In theory, this information could be used for constraining molecular docking efforts. However, significant artefacts can be introduced by the inherent differences in transverse relaxation rates of ligand resonances. Therefore, the most significant advantage of the paramagnetic methods is that structural information relative to a fixed point on the protein can be reliably obtained from the ligand spectra. Previous paramagnetic NMR methodologies for determining protein-ligand structures have been limited to metalloproteins with an intrinsic metal binding site[97] or have used a lanthanide tag that can only be placed at the N-terminus of a protein.[99] In contrast, the CLaNP-5 tag can be used on non-metalloproteins and can be placed at a variety of sites on the protein, provided they are sufficiently rigid. It should be noted that in our hands, the effect of the tag on the stability of the protein is variable. In some cases, but certainly not all, the protein more easily precipitates, perhaps due to the partly hydrophobic nature of the cyclen ring system. In general we design several extra double Cys mutants and

use the most stable ones. The tag is attached via disulphide bridges, which means that the probe can be lost. The rate of dissociation is variable (days-weeks), so it is advised to prepare the samples freshly or store them at -80ºC.
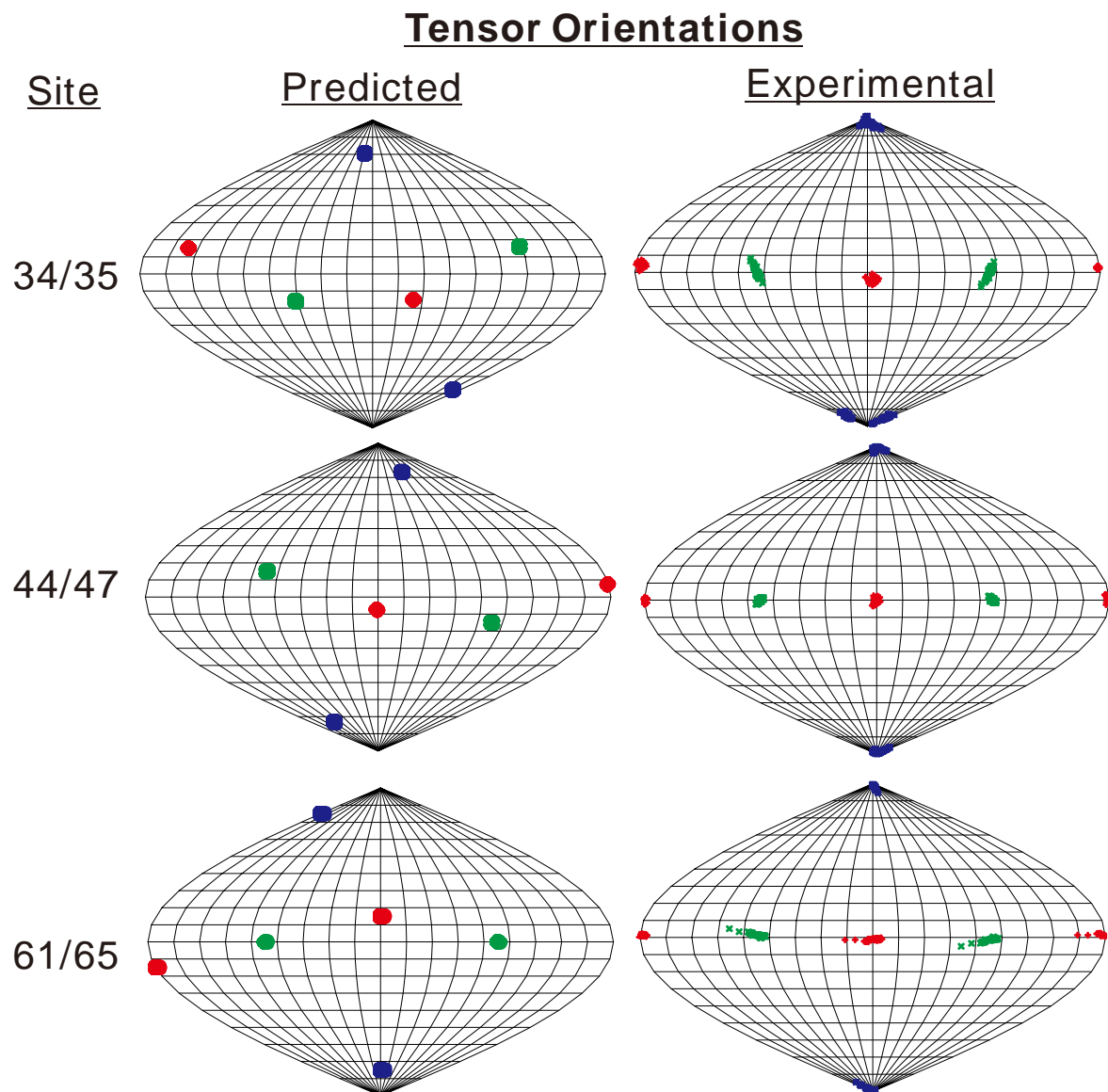
Figure 4.9: Sanson-Flamsteed projections of the predicted and experimental tensors. The orientations of the experimental tensors were translated to the axes of the Cartesian coordinates for an easier comparison with the predicted tensors. Blue, green and red colors represent z, y and x axes. The errors in the experimental tensors were calculated by randomly excluding 10% of the data. The orientations of the predicted tensors were modeled based on the protein structure. The intervals between the grid lines are 20 degrees.

The principle of the proposed technique is based on determination of the ligand position with respect to the paramagnetic $\Delta\chi$-tensor. In order to predict the $\Delta\chi$-tensor positions, the protein structure must be available. The results presented in this study have demonstrated that, with the predicted $\Delta\chi$-tensor parameters, it is possible to identify the ligand binding site. It is not necessary to have protein resonance assignments in order to predict the $\Delta\chi$-tensor position and therefore the approach can be applied to proteins that cannot be readily isotopically labeled. In principle, the approach can also be applied to proteins for which experimental structure information is not available, providing reliable structure prediction methods (such as homology modeling[193,194]) are applicable. For drug discovery, the potential binding site can be identified, which can accelerate optimization of hits to achieve higher affinity and greater biological activity even when the structure of the target is not available.

4.5   Conclusions

We have successfully determined the site and orientation of a small molecule ligand binding to a protein using ligand PCS and validated the results with an NOE structure. The use of the CLaNP-5 tag to induce the paramagnetic effects makes this approach suitable for non-metalloproteins. The results show that this strategy can identify the ligand binding site better than chemical shift perturbations. Comparison of the PCS structures from predicted and experimentally determined tensors demonstrates that the predicted tensor positions are sufficient for coarse definition of the binding site and it is therefore not necessary to experimentally optimize the tensor position. This PCS-based approach can be useful in early stages of fragment-based drug discovery to identify binding sites for proteins that are difficult to be enriched with isotopes, and in this way support optimization of early fragment hits.

5

An ensemble of rapidly interconverting orientations in electrostatic protein-peptide complexes characterized by NMR spectroscopy

Abstract

Protein complex formation involves an encounter state in which the proteins are associated in a non-specific manner and often stabilized by electrostatic interactions between charged surface patches. Such patches are thought to bind in many different orientations with similar affinity. To obtain experimental evidence for the dynamics in encounter complexes, a model was created using the electron transfer protein plastocyanin and short charged peptides. Three plastocyanins with distinct surface charge distributions were studied. The experimental results from chemical shift perturbations, paramagnetic relaxation enhancement (PRE) NMR and theoretical results from Monte Carlo simulations indicate the presence of multiple binding orientations that interconvert quickly and are dominated by long-range charge interactions. The PRE data also suggest the presence of highly transient orientations stabilized by short range interactions.

5.1 Introduction

According to the current models, formation of a specific protein complex is preceded by that of an encounter complex.[195] It is believed that in this state the partners assume multiple orientations to enhance the probability of finding the specific binding site.[196] Often, in the encounter complex electrostatic interactions dominate whereas the specific (final) state is stabilized by various short-range interactions. The assumed presence of multiple orientations in the encounter state is based on the theoretical notion of charged surface patches. Like velcro,[197] such patches can bind in many orientations with similar energy and thus all are assumed to be populated. The presence of multiple orientations and the dynamic exchange between them in the charge-driven encounter state is, however, not easy to demonstrate experimentally.

The aim of this study was to create a pure, charge-driven encounter state, and demonstrate the existence of a rapidly exchanging set of binding orientations. We chose to study the complex of plastocyanin (Pc) and short charged peptides (Lys$_4$), assuming that the interaction would be dominated by the strong positive charges of the peptides. The peptides are an artificial binding partner, so plastocyanin will not have an optimized binding site and a specific complex is unlikely to be formed.

Pc is a type-I blue copper protein involved in the electron transport process in oxygenic photosynthesis, functioning as an electron carrier between cytochrome *f* (Cyt *f*) of the b_6f complex and P700$^+$ of photosystem I (PSI). Structures are available for Pc from various plants and bacteria.[36,41–43,45,53,54,198] The C-terminal histidine that is a copper ligand is the electron entrance, located at the so-called 'northern' side of the protein, placed in a hydrophobic patch. Pc is acidic in higher plants[41,51,53,54,198] and green algae,[199–201] possessing two highly conserved negatively charged surface regions (acidic patches) formed by amino acids at positions 42–44 and 59–61 at the so-called 'eastern' side. The typical example of *Populus nigra* Pc (PoPc) is shown in Figure 5.1A. Compared to typical plant Pcs, the structure of Pc from the fern *Dryopteris crassirhizoma* (DPc) conserves the same global structure (Figure 5.1B), but a large acidic arc extends to the north side surface near the hydrophobic patch, resulting in very distinct electrostatic properties.[43] In cyanobacteria, Pc can also be almost neutral[42,202] such as *Phormidium laminosum*[42] (Figure 5.1C), or basic, such as *Nostoc* sp. PCC 7119.[203–205]

Charged peptides have proved useful to study interacting sites in proteins and the corresponding structural changes in electron transfer proteins, including Pc, Cyt *f* and Cyt *c*.[44,146,147,206–209] Experimental evidence showed that positively charged polylysine peptides

interact with the clustered acidic residues on Pc and competitively inhibit electron transfer from Cyt *c* or Cyt *f* to Pc.[147,207] The resulting competitive inhibition was explained by neutralization of charges by the formation of the Pc-peptide complexes.[147] The experimental evidence has also shown that binding of polylysine peptides to Pc can subtly perturb the active site geometry and the redox potential.[44,147] Little information is, however, available on the binding interface and the underlying degree of dynamics in the interaction.

Paramagnetic relaxation enhancement (PRE) NMR spectroscopy has been used as a sensitive tool to detect lowly-populated intermediates in biomolecular complexes.[112,121] The large magnetic moment of the unpaired electron from the paramagnetic center causes relaxation of a nucleus in the vicinity, and this effect falls off very rapidly, being proportional to the inverse sixth power of the distance between the electron and the nucleus. TOAC (2,2,6,6-tetramethyl-N-oxyl-4-amino-4-carboxylic acid) has been shown to be useful for PRE NMR studies of protein-peptide interactions.[143] One of the advantages of TOAC over side chain attached spin labels is that TOAC can be directly incorporated into the peptide backbone in automated peptide synthesis. There has been growing interests in using TOAC in peptide-protein and peptide-nucleic acid interactions and in combination with other physical techniques, such as electron paramagnetic resonance (EPR), circular dichroism (CD), fluorescence, Fourier-transform infrared (FT-IR), NMR, and X-ray crystallography, to understand molecular interactions.[210]

In this study, the transient complexes formed by tetralysine peptides and three different plastocyanins have been studied using chemical shift perturbation (CSP) analysis, PRE NMR, ensemble docking and Monte Carlo (MC) simulations. The CSP data corresponded well with the electrostatic MC docking calculations, clearly showing that the binding is dominated by charge interactions. The PRE data indicated that within the electrostatic ensemble, the peptides assume multiple orientations in a dynamic fashion. The PRE data also provide evidence for the presence of orientations that are slightly more favored than expected from pure electrostatic interactions, perhaps due to transient hydrogen bond formation with TOAC or weak hydrophobic interactions. Overall, the experimental and simulation results provide direct evidence for dynamics in an encounter complex dominated by electrostatic interactions.
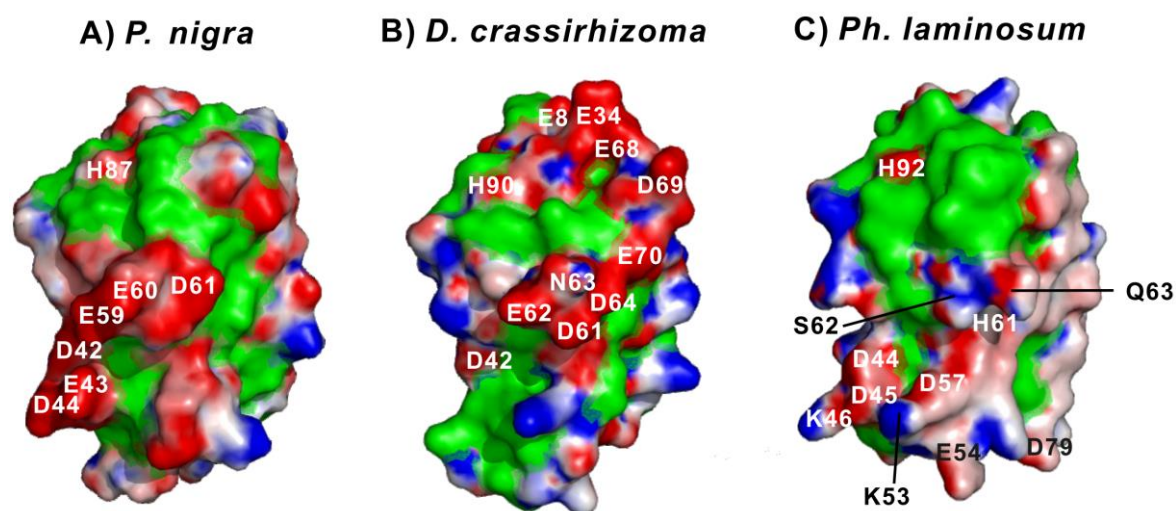
Figure 5.1. Electrostatic potential maps of Pc surface models (PDB entries 1TKW,[47] 1KDI[43] and 2Q5B). The surfaces are colored according to the electrostatic potential calculated by the program APBS[211] at an ionic strength of 10 mM, pH 6.5, 300 K to match the experimental conditions. The electrostatic potentials are colored and contoured from -8 (intense red) to + 8 kT/e (intense blue). Hydrophobic residues (A, V, I, L, F, P, Y and M) are colored in green. Several relevant residues are labeled. For PhPc (C) two additional acidic residues are labeled (D79 and E54). Pictures were generated with the program PyMOL.[169]

5.2   Materials and Methods

Peptide synthesis and preparation

Fmoc-TOAC-OH was purchased from Iris Biotech (Germany). Synthetic peptides AKKKK, KKKKA, TOAC-KKKK (XKKKK) and KKKK-TOAC (KKKKX) were prepared as described,[143] with N-terminal acetylation and C-terminal amidation. Peptides were checked on purity using rpHPLC and on integrity using MALDI-TOF mass spectrometry. The peptides were dissolved in 10 mM NaPi, pH 6.5. The fraction of paramagnetic peptide was checked by EPR and found to be close to 100%. The quantity of trifluoroacetic acid (TFA) in the sample was confirmed by ^{19}F-NMR with trifluorotoluene as the internal reference. The molar ratio of 5:1 for TFA: peptide was used to calculate the peptide concentration. The peptides were kindly provided by Dr. Jan Wouter Drijfhout (Leiden University Medical Center). Martin van Son (Leiden University) is acknowledged for the EPR measurements.

Protein expression and purification

General procedure:

^{15}N-enriched M9 minimal media was prepared as described previously.[212] For PoPc and PhPc, copper was excluded during bacterial growth. For additional ^{13}C labeling the minimal medium was supplemented with 2 g/L ^{13}C-Glucose. Cells were harvested by centrifugation and lyzed with a French pressure cell (Stansted Fluid Power Ltd.) in the presence of 1 mg lysozyme, 3.75 mg DNase, 1 mM PMSF and $ZnCl_2$ (100 µM for PoPc and DPc, 5 mM for PhPc). For PoPc and DPc, an additional 250 µM $ZnCl_2$ was added after passing through the French press. Cell debris was removed by centrifugation at 7,000× g for 25 min and membranes were removed by ultracentrifugation at 25,000× g for 1 h. All columns used for purification were purchased from GE Healthcare Biosciences. PoPc and DPc concentrations were determined using the Bradford assay (Bio-Rad) with bovine serum albumin as the standard. Pc was considered pure when the protein migrated as a single band on SDS-PAGE.

PoPc: The PoPc gene from plasmid pETPc[47] was subcloned to pET28 with an additional glycine residue at the N-terminus. ^{15}N-labeled PoPc was essentially produced as described[47] with the following modification: the protein was expressed in *E. coli* (Rosetta 2) in M9 minimal medium. Protein production was induced by adding IPTG to a final of 0.5 mM IPTG. Incubation was continued at 16 °C overnight. The protein was purified using 3× 5 mL HiTrap-DEAE FF ion-exchange column chromatography in 20 mM sodium phosphate, pH 7.0. The protein was eluted with a gradient of 0-500 mM NaCl. Fractions containing PoPc were concentrated and purified by a Superdex G-75 size exclusion chromatography in 20 mM sodium phosphate, pH 6.8, 100 mM NaCl. The yield of pure protein was 1.5 mg/L of culture for ^{15}N-PoPc and 0.75 mg/L of culture for ^{15}N,^{13}C-PoPc. Anneloes Blok is acknowledged for the construction of the plasmid and Dr. Monika Timmer for providing ^{15}N,^{13}C-PoPc.

DPc: ^{15}N- and ^{15}N/^{13}C-labelled recombinant DPc containing zinc was produced in *E. coli* BL21(DE3) and purified as described before,[46] with the following modifications: All copper salts were replaced by $ZnCl_2$ during purification. The protein was purified using 3× 5 mL HiTrap-Q HP ion-exchange column chromatography in 10 mM sodium phosphate, pH 5.8 at 4 °C. The impurities were eluted with a gradient of 0-100 mM NaCl at 4 mL/min and the Pc protein was eluted at 100 mM NaCl at 0.5 mL/min. Then, size exclusion chromatography with Superdex G-75 was performed in 10 mM sodium phosphate, pH 6.5, 100 mM NaCl. The yield of pure protein was 149 mg/L of culture for ^{15}N-DPc and 19 mg/L of culture for ^{15}N,^{13}C-DPc.

PhPc: Uniformly ^{15}N-enriched PhPc was produced without the leader peptide and purified as described,[213] with the following modifications: After cell lysis and ultracentrifugation, the supernatant was dialyzed against 0.5 mM $ZnCl_2$, 5 mM Tris-HCl pH 7.5, overnight at 4 °C. Pc concentrations were determined using ε_{280} = 5.00 cm^{-1} mM^{-1}. The yield of pure protein was 3.5 mg/L of culture for ^{15}N-PhPc. Dr. Monika Timmer is acknowledged for providing ^{15}N-PhPc.

NMR measurements

All Pcs were concentrated by ultra filtration (Amicon, MW-cutoff 3 kDa). The samples were in 10 mM sodium phosphate, pH 6.5, 6% D_2O. For peptide titrations, the protein concentrations were 200 μM for ^{15}N-DPc(Zn), 200 μM for ^{15}N-PhPc(Zn) and 110 μM for ^{15}N-PoPc(Zn). The samples for fern Pc and poplar Pc backbone assignment consisted of 2.4 mM and 0.25 mM $^{13}C/^{15}N$ labeled protein, respectively. Peptide solutions were prepared in 10 mM sodium phosphate, pH 6.5. All NMR spectra were recorded at 300 K on a Bruker AVIII600 spectrometer equipped with a triple-resonance TXI-Z-GRAD cryoprobe, or a Bruker 600MHz Avance DRX spectrometer equipped with a 5 mm TCI cryoprobe. Data were processed with TopSpin and analyzed in SPARKY. Resonances in the HSQC spectra of DPc and PoPc were assigned using a 3D HNCACB experiment. The side chain resonance assignments of PoPc were taken from PoPc(Cd).[47] NMR assignments have been deposited to the BMRB, entry codes 19236 (DPc) and 19247 (PoPc), and are available in Appendices D and E.

PRE analysis

The paramagnetic XKKKK and KKKKX peptides were added into ^{15}N-labelled Pc separately and [^{1}H,^{15}N]-HSQC spectra were recorded. Each paramagnetic peptide was added to Pc at a peptide/Pc molar ratio of 0.5 for DPc, and 1 for PoPc and PhPc. Under these conditions, the fractions of bound Pc were 14% for DPc, and 35% for PoPc. The diamagnetic spectra were recorded by reducing the peptides with 5 equivalents of sodium ascorbate. The PREs were determined according to the procedure of Battiste and Wagner.[214] The intensity ratio I_{para}/I_{dia} of the Pc resonances in the presence of XKKKK or KKKKX was normalized by dividing them by the average value of the ten largest I_{para}/I_{dia} values. The scaling factors for each Pc-peptide are 0.92, 0.87, 0.93, 1.07, 0.94 and 0.95 for PoPc-KKKKX, PoPc-XKKKK, DPc-KKKKX, DPc-XKKKK, PhPc-KKKX and PhPc-XKKKK, respectively. The PRE (R_2^{para}) values were calculated according to Equation 1.2. The transverse relaxation

rates in the diamagnetic sample (R_2^{dia}) were calculated from the line width at half height obtained from a Lorentzian peak fit in the direct dimension, by using SPARKY.[164] The symbol t denotes the time for transverse relaxation during the pulse sequence (9 ms).

Calculations of dissociation constants and bound ligand fractions

Peptide binding was observed via the changes of protein resonances in the [^{1}H,^{15}N]-HSQC spectrum upon titration with the peptide. CSP analysis was carried out as described before.[215] The dissociation constants (K_D) for PoPc were determined using a two-parameter non-linear regression curve fitting based on a one-site binding model as described previously.[166] The fraction of bound ligand was calculated using the dissociation constant. For DPc, resonance overlap was observed for T79/E25 and A23/F12 during the titrations. These four residues were excluded from the K_D calculation. The peptide:DPc interaction was modeled with two independent binding sites (Scheme 5.1).

$$
\begin{array}{ccc}
 & K_{D1} & \\
P + L \rightleftharpoons & & (PL)_1 \\
K_{D2} \Updownarrow & & \Updownarrow +L \\
 & & K_{D2} \\
(PL)_2 \rightleftharpoons & & PLL \\
 & +L & \\
 & K_{D1} &
\end{array}
\qquad \text{(Scheme 5.1)}
$$
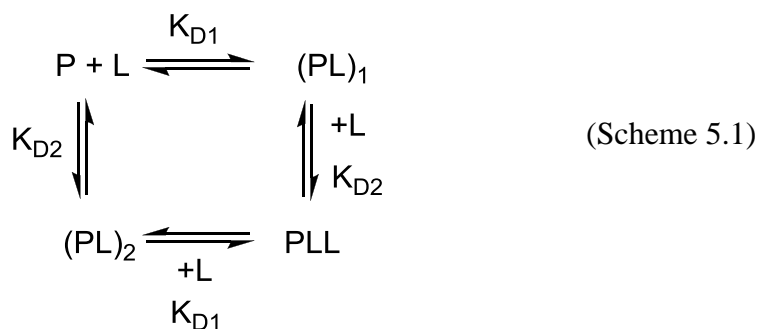
Here P and L are the free protein and the free peptide, respectively. (PL)$_1$ and (PL)$_2$ are the 1:1 complexes formed by peptide binding to sites 1 and 2 on DPc, respectively. PLL is the protein with two peptides bound. K_{D1} and K_{D2} are the dissociation constants for sites 1 and 2, respectively. The binding curves were simulated numerically with varying values for K_{D1}, K_{D2} and the $\Delta\delta$ at 100% bound Pc using Microsoft Excel.

Ensemble docking

For DPc, PhPc and PoPc the PDB entries 1KDI[43], 2Q5B and 1TKW model 1[47] were used, respectively. The PREs were converted into distances for structure calculations as described previously.[215] τ_c was taken to be 5.54 ns for DPc, 5.14 ns for PoPc, and 5.93 ns for PhPc, on the basis of the HYDRONMR[216] prediction of the rotational correlation time for each Pc. For each peak, R_2 was estimated from the width at half-height ($\Delta\nu_{1/2}$) of a Lorentzian fit in the proton dimension by using $R_2 = \pi\Delta\nu_{1/2}$. PRE values were calculated after normalization of the I_{para}/I_{dia} ratios and extrapolated to 100% bound by dividing the values by their bound

fractions (35% for PoPc and 14 % for DPc). Three classes of PRE restraints were included in the calculations.[143] 1) For the amides of which the resonances disappear in the paramagnetic spectrum, an upper limit for I_{para} was estimated from the standard deviation of the noise level of the spectrum. The upper bound PRE (R_2^{para}) value was set to 500 s^{-1} and the distance to 9 Å. 2) For residues with I_{para}/I_{dia} >0.85, the lower bound distance was set to 15 Å. 3) For residues with I_{para}/I_{dia} between 0.1 and 0.85, the distances (r) calculated according to a previously described equation[215] were used with upper and lower bounds of r ± 0.1 Å. Violations were defined as the absolute differences between the calculated distance and the experimental distance including the corresponding upper and lower bound margins for the three classes. An additional restraint ensures that the TOAC nitroxy oxygen atom and the Pc center of mass are at a distance between 10 and 30 Å. The structure calculations were done in XPLOR-NIH.[178] The accessible surface area (ASA) of each amide proton was calculated with a Python-based implementation of the Shrake-Rupley algorithm.[217]

Monte Carlo (MC) simulation

The peptide coordinates of XKKKK and KKKKX were generated from PRODRG server[153] and the conformations were optimized in Swiss PDB-Viewer[218] to separate the charges as far as possible. For DPc, PhPc and PoPc the PDB entries 1KDI[43], 2Q5B and 1TKW model 1[47] were used, respectively. The structure preparation and the Monte Carlo simulation[219] was performed as described.[180,220] The electrostatic potential was calculated with APBS[211] for an ionic strength of 0.01 M and a temperature of 300 K to match the experimental conditions. An ensemble of 2000 peptide orientations, randomly selected from the entire run of 2.2×10^6 saved structures, was considered for the calculations. The averaged distances were derived from the ensemble and compared to the experimental distances. Johannes M. Foerster (University of Bayreuth, Germany) is acknowledged for performing the MC simulations.

5.3 Results

Backbone assignments

In order to study the three Pcs with NMR, the proteins were isotopically labeled with ^{15}N for the PRE measurements, and ^{15}N/^{13}C for the resonance assignments. To eliminate the paramagnetic effect of Cu^{2+}, Zn-substituted Pc was used. For DPc and PoPc, backbone amide resonances were assigned using the HNCACB experiments on ^{13}C/^{15}N-labeled proteins. The assignments of Cu(I)-DPc (BMRB code 7370)[46] and Cu(I)-PoPc (BMRB code

4019) were used as the starting points. The data of backbone assignments (H, N, C_{α}, C_{β}) have been deposited to BMRB under the codes 19236 (DPc) and 19247 (PoPc) and are provided in Appendices D and E. The assignments of Zn-substituted PhPc were kindly provided by Dr. Sandra Scanu (Leiden University). For DPc, the resonance of S92 was not found in the spectra. For PoPc, some residues close to N-terminus have double peaks. The double resonances exist for I1, D2, V3, S20, I21, S22, P23, G24, E25, K26, I27, V28, K30, M57, T69, F70, E71, V72, L74 and G78. Similar observations were described for Cd-PoPc.[47] The double signals have been attributed to partial processing of the N-terminus methionine in the bacterial cytoplasm, as these residues are located near the N-terminus in the three-dimensional structure of the protein.[47]

Chemical shift perturbations

To study the interaction of Pc with lysine peptides, four types of peptides were used. For the PRE experiment described below, a TOAC residue (X) was introduced at the N or C terminus (XKKKK and KKKKX). As controls for the introduction of TOAC, AKKKK and KKKKA were also used. First, the interaction of these peptides with the three Pc variants was studied using chemical shift perturbation (CSP) analysis.

Each ^{15}N Pc was titrated with the four peptides individually in a low ionic strength buffer (I = 10 mM), and [^{1}H,^{15}N]-HSQC spectra were acquired at each titration point. For these studies TOAC was reduced to eliminate its paramagnetic effects. Addition of the peptides gave rise to small CSPs in the [^{1}H,^{15}N]-HSQC spectra of all Pcs, with maximum observed average shifts, $|\Delta\delta_{ave}|$ of, 0.07 ppm for PoPc, 0.05 ppm for DPc, and 0.01 ppm for PhPc (Figure 5.2). Single, averaged resonances were observed for all amides, indicating fast exchange between the free and bound Pc on the NMR time scale. The binding maps, obtained by coloring the protein residues according to the size of CSP, show similar patterns for different peptides for the same Pc (Figure 5.2 for KKKKX and Figure F1 for the other peptides). The similar patterns observed for KKKKX and XKKKK indicate that the CSPs are dominated by interaction with the four lysines. The binding maps of AKKKK and KKKKA were also similar to those of XKKKK and KKKKX, indicating no significant effect of TOAC on the peptide binding (Figure F1).

In PoPc and DPc, most CSPs occurred around the regions of the acidic patches, in agreement with the assumption that the positively charged peptides interact with the acidic residues of Pc.[145,147] The largest CSPs for PoPc occurred for residues D44, S45, D51, I55, and Gln88. Among these residues, D44 belongs to the acidic patch. For DPc, the largest CSPs occurred

for residues V3 and E8. E8 is located at the acidic arc of the northern side. Although the observed CSPs are very small for PhPc, still similar effects are observed from both TOAC peptides (Figure 5.2C). The small perturbations of the copper ligand residues (H37, C84, H87 and M92 for PoPc; H37, C87, H90 and M95 for DPc; H39, C89, H92 and M97 for PhPc) indicate that the copper site is not the main binding site of the peptides. Similar magnitudes of perturbations and binding maps caused by a tetralysine peptide (without an additional TOAC) were observed for Pc from the seed plant *Silene pratensis*.[221]
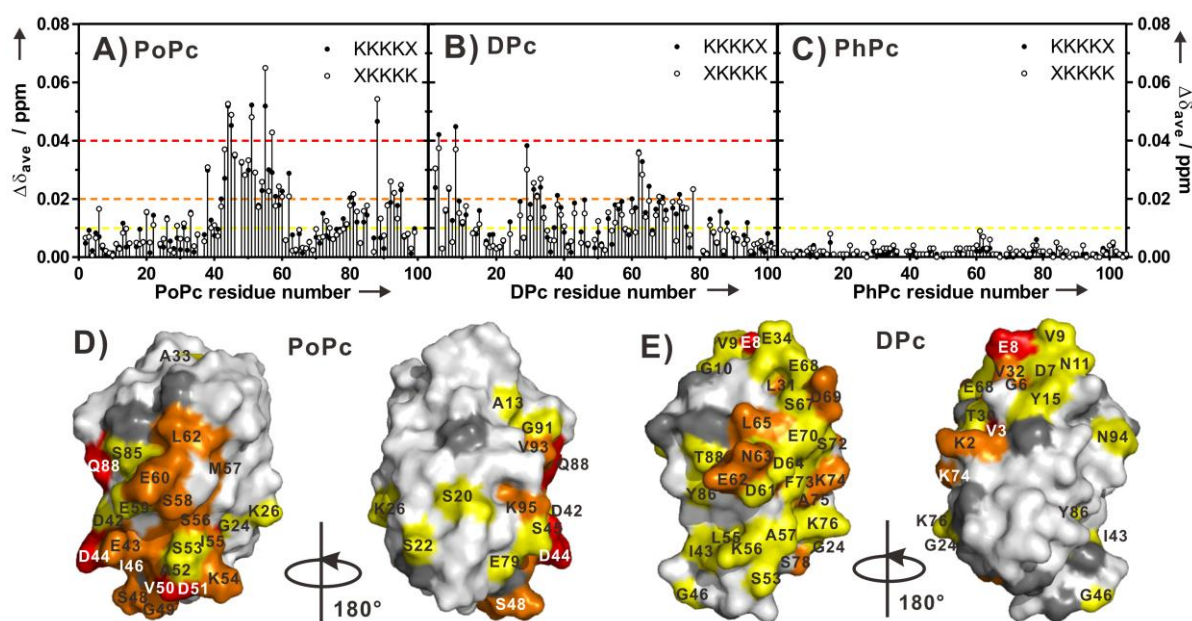
Figure 5.2: (A-C) Plots of NMR chemical shift perturbations measured for Pc backbone amides in the presence of TOAC-containing peptides. Extrapolated values (to 100% bound) for PoPc and DPc, and observed values for PhPc are shown. (D-E) CSPs (extrapolated to 100% bound, bound fractions see Table 1) mapped on the protein surfaces from the binding of KKKKX to PoPc (panel D, PDB entry 1TKW[47]) and DPc (panel E, PDB entry 1KDI[43]). Color representations: red, $\Delta\delta_{ave} \geq 0.04$ ppm; orange, $0.04 > \Delta\delta_{ave} \geq 0.02$ ppm; yellow, $0.02 > \Delta\delta_{ave} \geq 0.01$ ppm; white, $\Delta\delta_{ave} < 0.01$ ppm; grey, no data or overlapping resonances. The binding maps for the other peptides are shown in Figure F1 in Appendix F.

Binding constants were obtained by fitting the chemical shift perturbation curves for the most affected residues (Figure 5.3, Figures F2 and F3 and Table 5.1). For PhPc, the magnitudes of observed CSPs were too small ($|\Delta\delta_H| \leq 0.01$ ppm) to determine a dissociation constant.
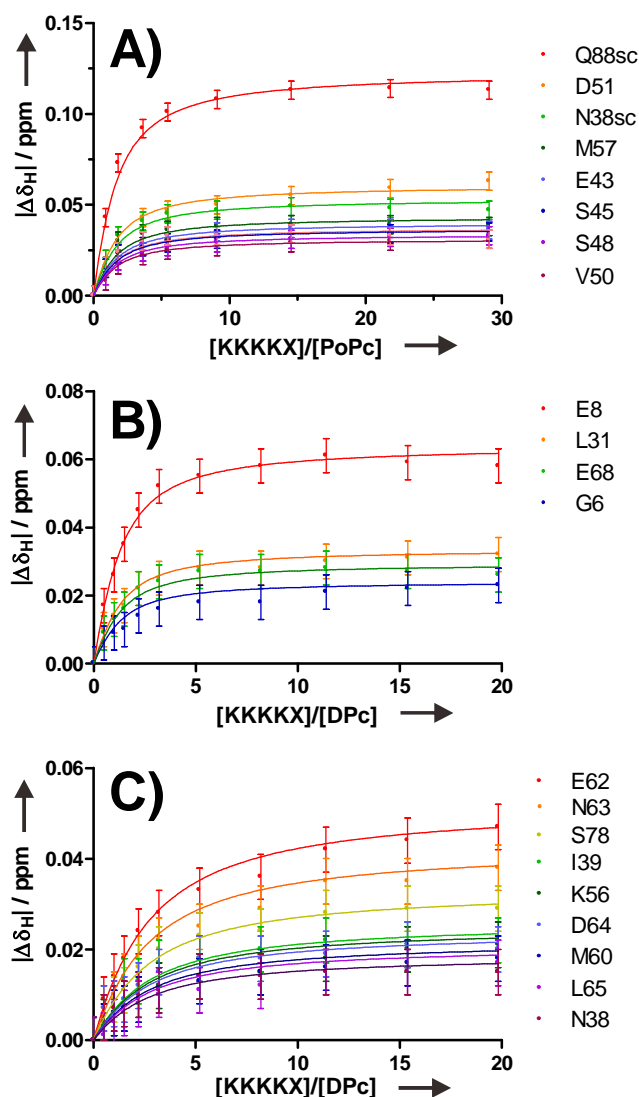
Figure 5.3: Chemical shift changes of selected Pcs resonances as a function of increasing [peptide]/[Pc]. The dissociation constants of the corresponding peptides (Table) were obtained by simultaneous fitting to a 1:1 binding model for PoPc (solid lines) and by simulation for 2-site binding for DPc. Error bars represent ± 0.005 ppm. (A) KKKKX with PoPc; (B) KKKKX with DPc, strong-binding residues; (C) KKKKX with DPc, weak-binding residues.

The binding curves for PoPc fitted well to a single binding site model (Figure 5.3A). Interestingly, there are two types of dissociation constants observed in DPc titrations. The residues that show stronger binding (lower K_D) are clustered on the northern side of DPc (Figure 5.2E and Figure F3B). This might be due to the unusual surface charge distribution of DPc compared with other plant Pcs. It is possible that there is internal competition between the two binding sites for the peptides. Clearly the 1:1 binding model is not appropriate to explain the observation. Therefore, a two-site binding model was used to obtain the K_D values for DPc (Figure 5.3B and C, Figure F2B-F2C, Figure F3C-F3F). For most peptides the K_D values for the same Pc are similar, indicating that the TOAC caused no significant changes in the affinity of the peptides for Pc. Only KKKKX has a somewhat lower K_D for PoPc than KKKKA, but the difference is within the error margins.

Table 5.1: Dissociation constants of the complexes formed between Pc(Zn) and tetralysine peptides and their calculated bound fractions (fr.) at the end point of the titrations.

Pc	KKKKA		KKKKX		AKKKK		XKKKK	
	K_D (μM)	fr.	K_D (μM)	fr.	K_D (μM)	fr.	K_D(μM)	fr.
PoPc	150±40	0.95	90±30	0.97	110±20	0.97	130±40	0.96
DPc (strong)	110±20	0.97	110±20	0.97	110±20	0.98	110±20	0.96
DPc (weak)	300±40	0.91	300±50	0.90	340±40	0.94	300±100	0.94

Paramagnetic relaxation enhancements

The paramagnetic TOAC was introduced to determine whether the bound peptide possesses a single, well-defined orientation or samples several orientations. If the peptide orientation is well-defined, the strong distance dependence of the PRE should result in highly localized effects. The position of the TOAC molecule was selected at the N- and C-terminal of the tetralysine peptide, in order to interfere minimally with binding.[143] The attached spin labels were thus expected to yield PRE of nuclei on nearby Pc residues. If the peptides bind in a specific orientation, the N- and C-terminal TOAC should generate different PRE patterns. PREs were observed for some residues, as shown in Figure 5.4. Binding of these peptides to the three Pcs is in the fast-exchange regime, so the observed PRE is a weighted average of free Pc (no PRE) and bound Pc. By dividing the observed PRE by the bound fractions calculated from K_D, the PRE for 100% bound Pc is obtained. For DPc, the weak-binding K_D values were used, because most residues showed weak binding.
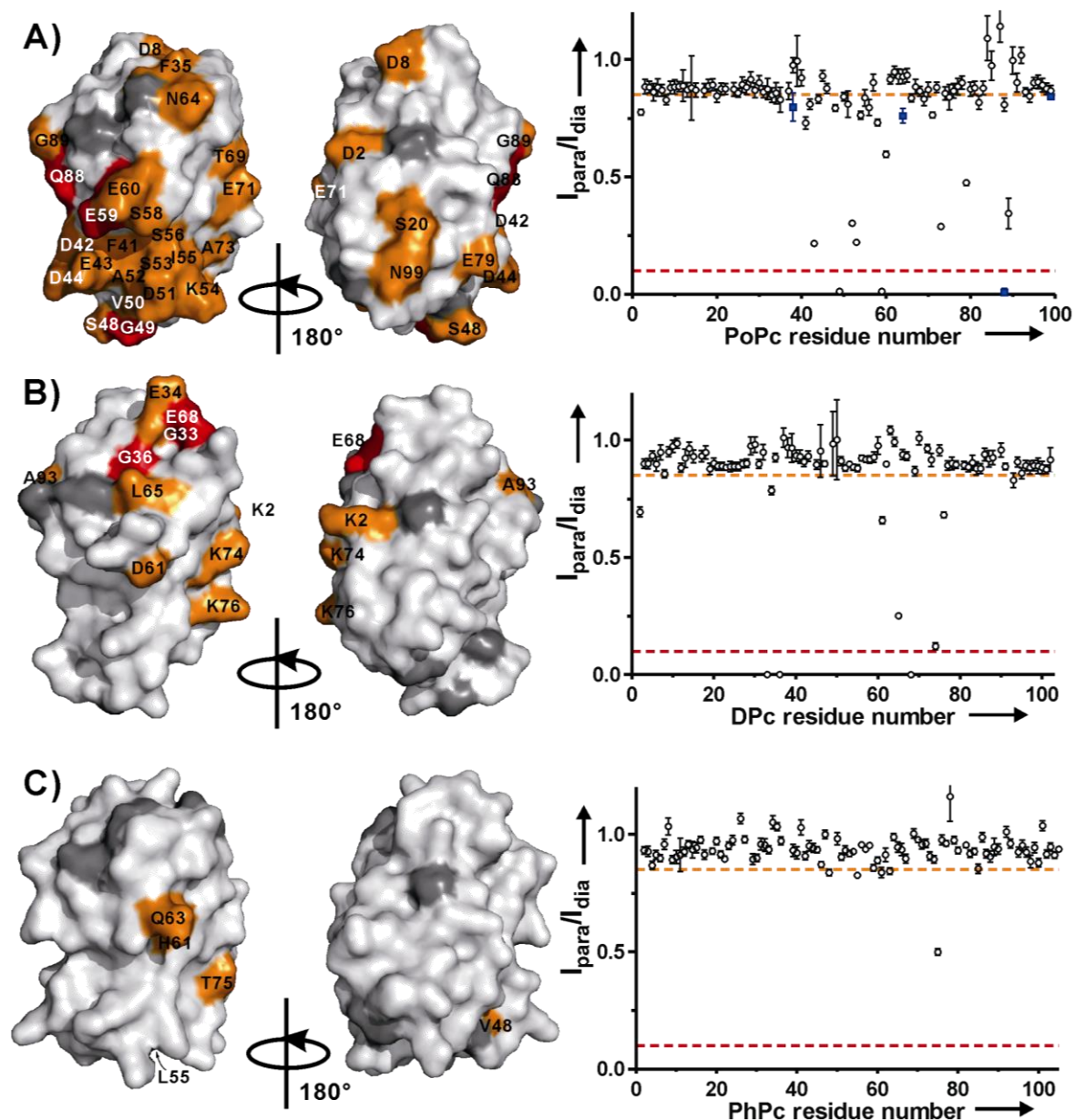
Figure 5.4: PRE effects in Pc-KKKKX complexes. The paramagnetic peptide was added to Pc at a peptide/Pc molar ratio of 0.5 for DPc, and 1 for PoPc and PhPc, resulting in fractions of bound Pc of 14% for DPc and 35% for PoPc. The bound fraction for PhPc is unknown but expected to be very small. Left: PRE maps of PoPc (A, PDB entry 1TKW[47]), DPc (B, PDB entry 1KDI[43]) and PhPc (C, PDB entry 2Q5B) bound to KKKKX peptide, color-coded on surface models of Pc: red, $I_{para}/I_{dia} < 0.1$; orange, $0.1 \leq I_{para}/I_{dia} < 0.85$; white, $I_{para}/I_{dia} \geq 0.85$; grey, prolines, unassigned, and overlapping resonances. Right: relative [^{1}H,^{15}N]-HSQC intensities of backbone amide of PoPc (A, including side chains which are shown in blue squares), DPc(B) and PhPc(C) in the complex with TOAC-containing peptides. The dashed horizontal lines indicate $I_{para}/I_{dia} = 0.85$ (orange lines) and 0.1 (red lines). The error bars denote 2× standard deviations, derived from spectral noise levels using standard error propagation procedures. For most data points, the error bars are within the symbol.

For PoPc binding to KKKKX, the residues that were broadened beyond detection were G49, E59, and the side chain of Q88. For PoPc binding with XKKKK, an additional residue (E43) was completely broadened. These residues are located on the same side as the acidic patches, which include E43 and E59. Many residues located around the acidic patch also experience PRE in different magnitudes. This observation indicates that the binding site(s) of the peptides on Pc is not restricted on the acidic patch residues only, but also extends to other polar or charged residues around this region and even to the hydrophobic patch, including some positively residues such as K26, K54, K66 (I_{para}/I_{dia} ratio 0.6-0.84). This observation suggests that the peptides sample a large area of the protein surface, and it demonstrates the superior sensitivity of PRE for transient interactions.

For DPc with KKKKX and XKKKK, three residues disappeared from the spectra: G33, G36 and E68. Two other acidic residues (E34, D69) are broadened but still visible in the spectra (I_{para}/I_{dia} ratio 0.59-0.82). These five residues are close together on the acidic arc at the northern side of DPc, indicating the cluster of negative charges on the protein attracted the peptides by electrostatic interactions.

For PhPc, only one residue (T75) had a clearly significant PRE (I_{para}/I_{dia} ratio 0.5) under the experimental conditions (peptide:protein ratio 1:1). The I_{para}/I_{dia} ratios of V48, L55, H61 were and Q63 and 0.84, 0.83, 0.84 and 0.83. These values are close to the defined threshold for unaffected residues (I_{para}/I_{dia} ratio 0.85).

The PRE effects of tetralysine peptides on DPc are smaller than on PoPc in general. This is due to a smaller bound fraction. The PRE maps of DPc showed much less effects than the CSP maps of DPc. The residues with largest CSP are not those with largest PRE, probably because CSP monitors the effects from all atoms within the peptides, whereas PRE indicates the effects from the paramagnetic center only.

It is interesting to note that strongly affected residues have unaffected neighbors. One example is A73 of PoPc, which is affected by PRE, while the neighboring V72 and L74 are not. Similarly, A75 of DPc, located in between the residues with PREs, K74 and K76, remains unaffected. Another example is seen for D61 and E62 of DPc, both located on the acidic arc. D61 is affected but E62 unaffected. These findings suggest highly localized effects and will be discussed in more detail later.

Ensemble docking

Visualization of the encounter state on the basis of PRE data can be carried out quantitatively using the ensemble docking approach.[112] Calculations were performed using 1

to 15 copies of a pseudoatom that represents the paramagnetic center. The experimental PREs were converted into distances for ensemble docking. For DPc, the K_D values used here are the low affinity ones, as most residues belong to the low-affinity group. The high-affinity residues were completely broadened and therefore their target distance ranges is the same using either K_D values. Violations were defined as the absolute differences between the distance back-calculated from the entire ensemble (using r^{-6} averaging) and the experimental distance. Figure 5.5 shows the results of ensemble docking for KKKKX binding to PoPc and DPc with increasing ensemble size. Large distance violations occurred when using a single representation of the paramagnetic center (Figure 5.5, N=1), indicating that multiple orientations are required to describe the data. As a result of increasing degrees of freedom, the distance violations are reduced with increasingly larger ensembles. For PoPc (Figure 5.5A), no significant reduction of violation occurred at N≥8. For DPc, the violation curve flattened at N=5 (Figure 5.5B).
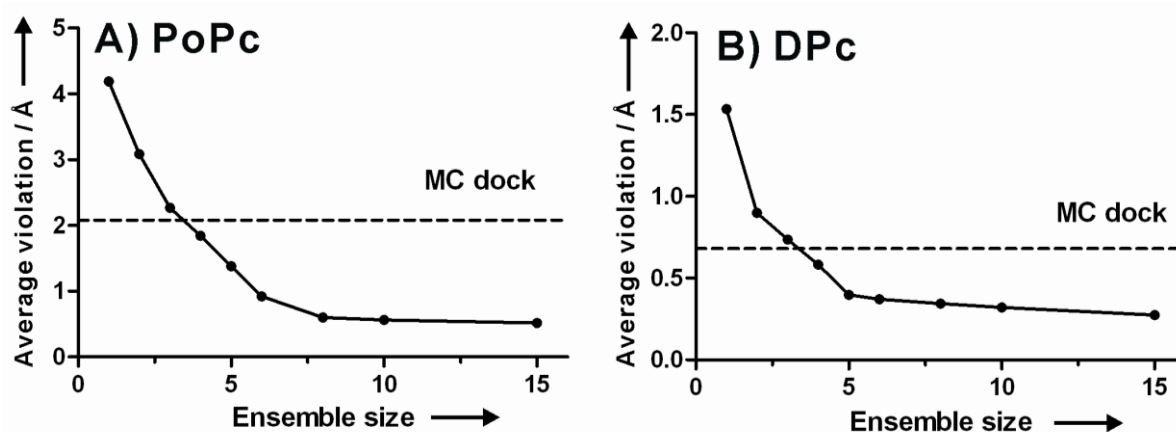
Figure 5.5: Averaged distance violations against number of paramagnetic pseudoatoms (N=1-6,8,10,15) in the ensemble docking. (A) KKKKX-PoPc, (B) KKKKX-DPc. The dashed horizontal lines indicate the average violations calculated from MC dock.

The resulting ensembles for N=6 are shown in Figure 5.6. Most of the paramagnetic centers are located in well-defined positions and not in a 'cloud' of orientations. This correlates with the observation that some amides of Pc are strongly affected by PRE whereas others that are nearby are not. That can be explained by assuming that the paramagnetic center spends a short time very close to the affected amide. Most of the affected amides have a considerable accessible surface area (ASA), enabling a close contact with the TOAC. In general, no major differences are observed for XKKKK and KKKKX.
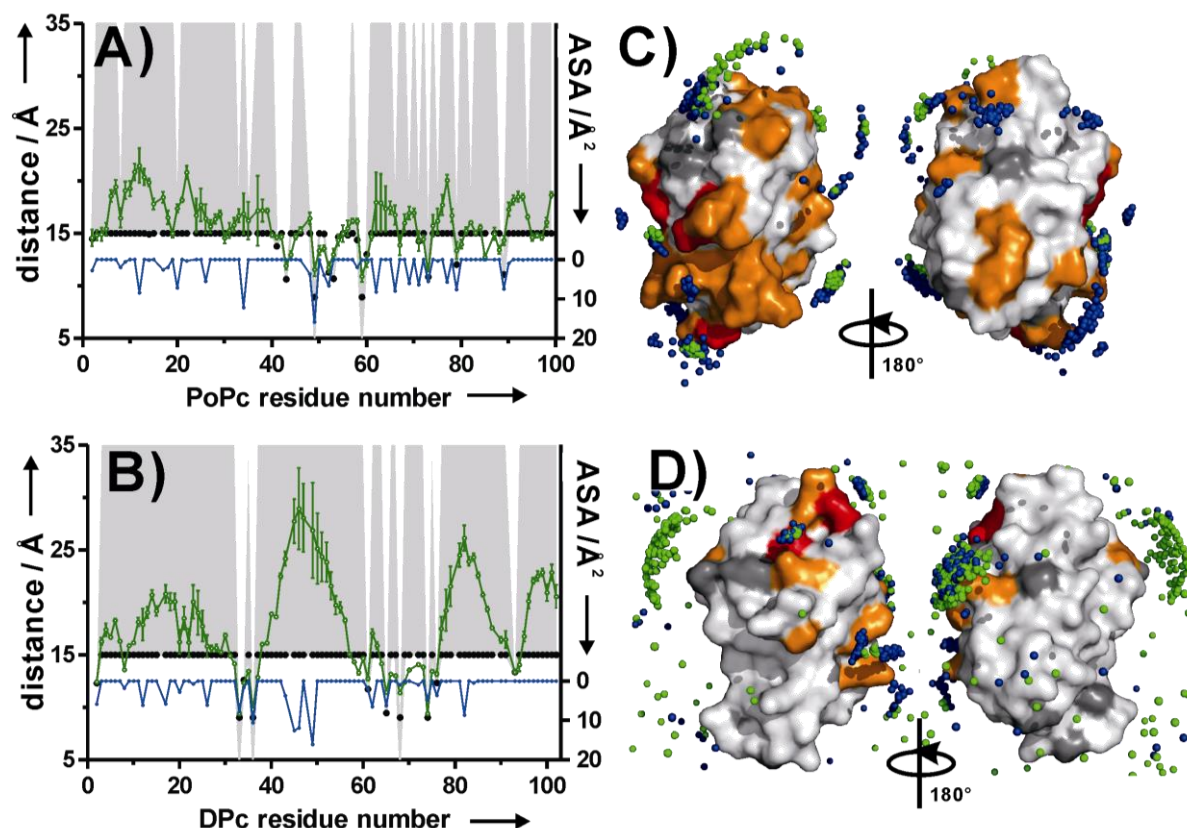
Figure 5.6: Ensemble docking. (A-B) Correlation of experimental distances (black dots) and back-calculated average distances (green dots with green connecting lines) from the ensemble docking (N=6) of KKKKX bound to PoPc (A) and DPc (B). The average distances from the 20 lowest-energy solutions of the PRE driven ensemble docking are shown as green circles connected by green lines with error bars representing the standard deviation. Right y axes indicate the accessible surface area (ASA) of each amide proton, shown in blue dots with blue connecting lines. Grey areas indicate the error margins of the experimental distances. (C-D) PRE-based ensemble docking results (N=6) of PoPc (C, 396 solutions for KKKKX and 594 for XKKKK) and DPc (D, 630 solutions for KKKKX and 360 for XKKKK). The paramagnetic centers from TOAC are shown as spheres, with KKKKX in green and XKKKK in blue. Protein surfaces are colored the same as in PRE maps (Figure 5.4).

Monte Carlo simulations

Previous studies have shown that many encounter complexes are predominantly driven by electrostatic forces,[196] although in some cases short-range hydrophobic interactions may also contribute.[220] Visualization of the encounter complex of Cyt *c* and Cyt *c* peroxidase was successfully achieved using PRE data and rigid-body MC simulations.[180] The results showed that this encounter complex is driven by electrostatic interactions. In MC simulations, one protein is docked to the other, guided by an electrostatic field and MC

sampling.[219] In this way, charge−charge interactions represent the only force that brings together the binding partners. MC simulations were performed for Pc-peptide complex, and the Boltzmann distribution of orientations of the peptide in complex with Pc was obtained. The paramagnetic centers of the peptides are shown as green (KKKKX) and blue (XKKKK) spheres around Pc in Figure 5.7.

The results for PoPc (Figure 5.7A) and DPc (Figure 5.7B) show that the peptides are located close to the acidic patches. For PhPc, the population is more randomly distributed, with a relatively higher density at the side of PhPc that is farthest from the hydrophobic patch (Figure 5.7C). Johannes M. Foester (University of Bayreuth, Germany) is acknowledged for performing the MC simulations.
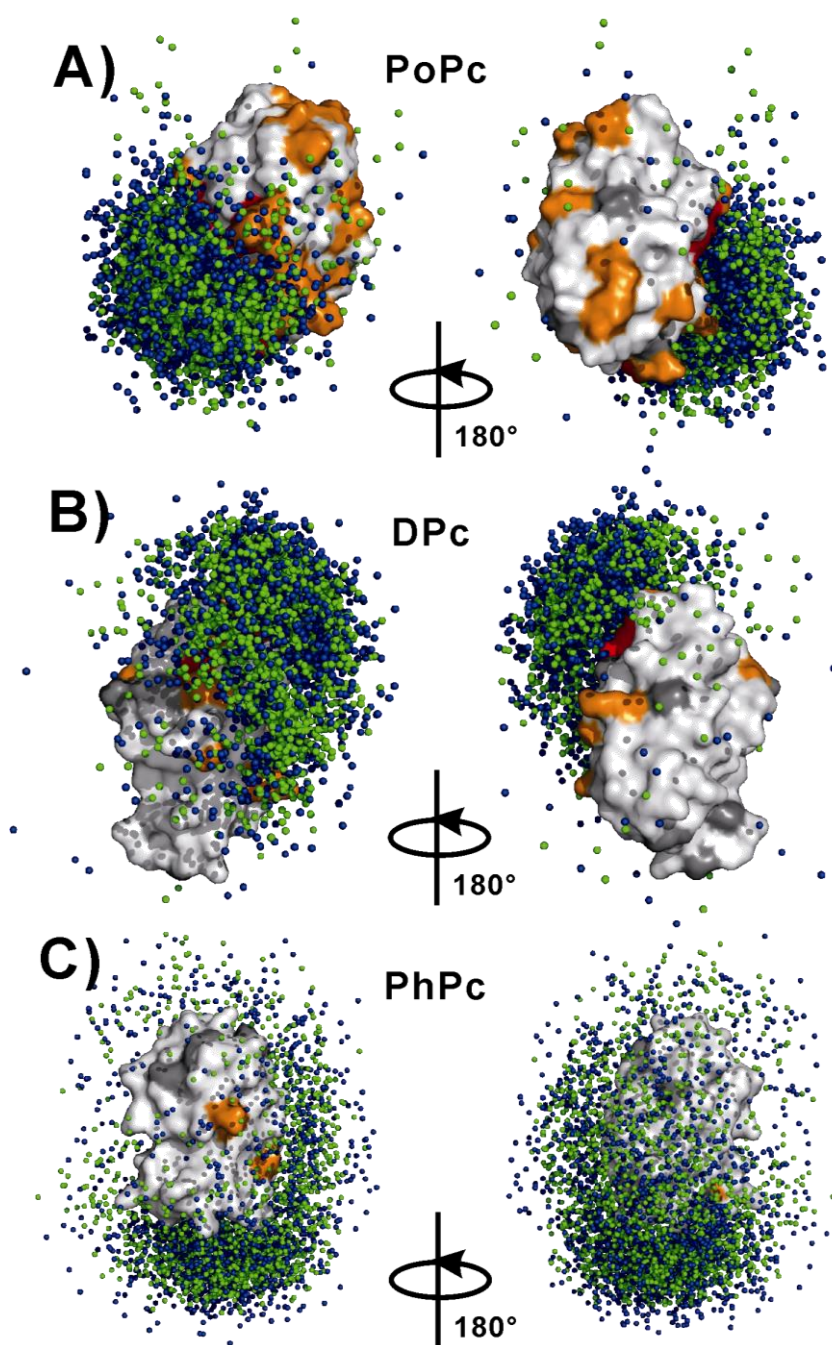
Figure 5.7: MC dock results showing 2000 solutions of KKKKX and XKKKK bound to PoPc (A), DPc (B) and PhPc (C). The paramagnetic centers of the peptides are shown as green (KKKKX) and blue (XKKKK) spheres. Protein surfaces are colored according to the PRE maps (Figure 5.4). The 2000 orientations in each ensemble were selected randomly for the entire MC dock solution set.

The distances from the nitroxy oxygen of the TOAC to the Pc amide hydrogens were measured and averaged (using r^{-6} averaging) for the 2000 MC ensemble, and the obtained distances were compared with the experimental values. The violations calculated for the MC dock ensemble are 2.08, 1.70, 0.68 and 0.56 for PoPc-KKKKX, PoPc-XKKKK, DPc-KKKKX and DPc-XKKKK, respectively. All violations are in the middle of the range of values shown in Figure 5.5 and Figure F5 (A and B). Thus, the MC dock ensemble does not fully agree with the PRE data. Figure 5.8 shows the back-calculated average distances for each Pc residue in comparison with the PRE derived distances. Although the MC dock clearly places the paramagnetic center close to the affected residues, the simulation underestimates the PRE for these residues.
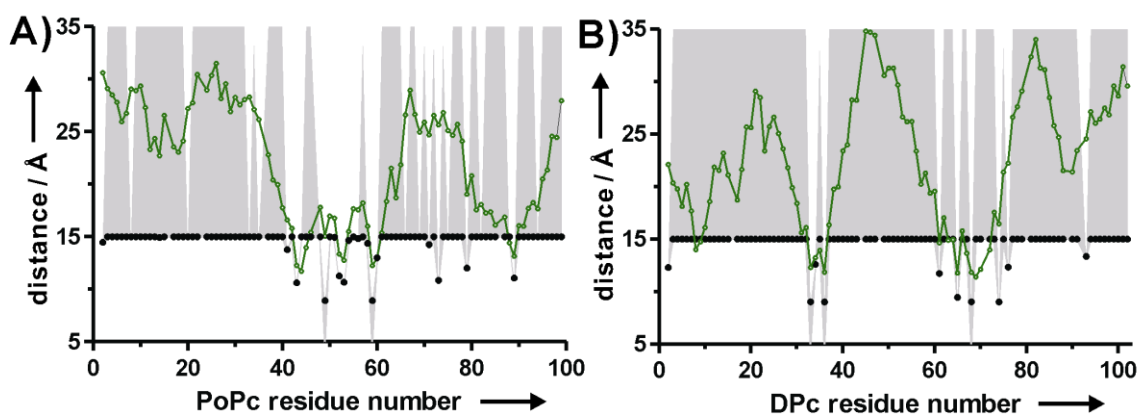
Figure 5.8: Comparison of experimental distances (dots) and back-calculated average distances (circles with connecting lines) between Pc amides and the 2000 ensembles of peptide paramagnetic oxygen atoms from MC simulations. (A) PoPc-KKKKX; (B) DPc-KKKKX.

Figure 5.9 shows the plots of electrostatic energy population distribution for all Pc-KKKKX complexes. PoPc (Figure 5.9A) and DPc (Figure 5.9B) have similar patterns. The highest population in DPc is at -6 kcal/mol, where as in PoPc is at -7 kcal/mol. For PhPc (Figure 5.9C), it is clear that electrostatic interaction is much weaker (highest population at -2 kcal/mol), indicating the electrostatic interaction is much weaker for PhPc. The histograms for Pc-XKKKK complexes are shown in Figure F5 in Appendix F. The highest population appears at -8, -7 and -2 kcal/mol for PoPc, DPc and PhPc, respectively.
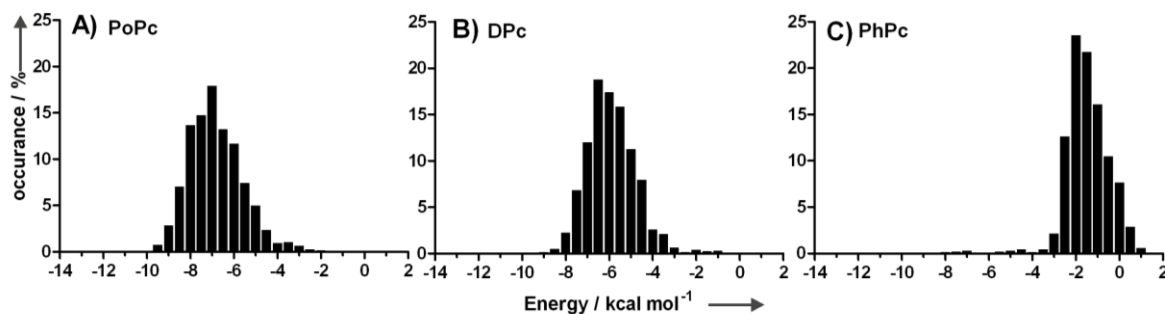
Figure 5.9: Histograms showing the electrostatic interaction energy distribution of 2000 structures randomly selected from the MC simulations. (A) PoPc-KKKKX, (B) DPc-KKKKX, (C) PhPc-KKKKX.

5.4   Discussion

The aim of the present work was to experimentally characterize the dynamics in encounter complexes. The rationale was to create a pure encounter complex by ensuring that electrostatics dominate the interactions. For this purpose the complexes formed by charged tetralysine peptides and three Pcs with distinct surface charge properties were studied. At pH 6.5, the net charge of PoPc, DPc and PhPc is -7, -5 and -1, respectively and the charge distributions differ markedly between these Pcs.

Previously, the interaction between the seed plant *Silene pratensis* Pc and lysine peptides of varying lengths was studied using circular dichroism, UV-Vis absorption, resonance Raman spectroscopy, and cyclic voltammetry. Minor changes in the geometry of the copper site were observed upon peptide binding.[44,147] The peptides also competitively inhibited electron transfer within Pc-Cyt *f*[207] and Pc-Cyt *c*.[147] Mutagenesis of Pc showed that the interaction and electron-transfer inhibition by lysine peptides decreased significantly as the net charge of the Pc negative patch decreased,[147] showing that charge interaction contributed to the binding. The authors proposed a specific and effective interaction between the positively charged peptides and the negative patches of Pc.[147] These studies monitored spectroscopic changes caused by peptide binding but could not directly observe the binding interface and the dynamics of the interaction.

Electrostatic interactions

To establish whether electrostatic interactions were the dominating interaction force, the interaction surface was mapped using CSPs and compared with electrostatically driven docking calculations. In PoPc and DPc, CSPs were largest in the acidic regions. The K_D values were about 100 μM for PoPc and 110 and 300 μM for the two binding sites on DPc.

In PhPc, peptide binding resulted in very small CSPs, suggesting a low affinity. No dissociation constant could be determined. These results are in good agreement with the MC simulations. The electrostatic ensembles match well with the CSP-derived binding maps for PoPc and DPc. The electrostatic interaction energies indicated that these two Pcs have a strong interaction, whereas for PhPc the affinity is quite weak. The data indicate that charge-driven binding is a good first description of the complexes.

Paramagnetic relaxation effects

To determine whether the peptides assume a single, well-defined orientation or sample multiple orientations, the paramagnetic amino acid TOAC was incorporated at the N- and C-terminus of tetralysine peptides. Control peptides with Ala instead of TOAC were used to assess the effect of TOAC incorporation on peptide binding to Pc. No significant difference between the binding affinities of TOAC- and Ala- tetralysine peptides was observed, indicating that TOAC has little influence on the thermodynamics of peptide binding.

In PoPc, the presence of TOAC caused PREs mainly in the neighborhood of the acidic patches as well as for some of the hydrophobic patch residues. CSPs were almost not observed in the hydrophobic patch, which suggests that the PREs for those residues represent peptide orientations that are sparsely populated. The PRE is highly sensitive for minor states in which the paramagnetic center is brought close to the nucleus. Apparently, transiently peptide-protein interactions that are not dominated by electrostatic forces are present. In DPc, the area affected by PREs is smaller and more localized than in PoPc. The largest PREs were detected around the top of the acidic arc, close to the copper site. The PRE and CSP maps are similar in this case. For PhPc, few PREs and CSPs were observed, in accord with the weak affinity for tetralysine peptides.

Dynamics in the complexes

It is believed that the overall size of the CSP is a measure for the degree of dynamics in a protein complex. Large CSPs are caused by a single, well-defined orientation in the complex, in which desolvation of the interface and multiple short-range interactions occur. Small CSPs indicate averaging of multiple orientations in the encounter state, with minimal desolvation. Small CSPs have been observed in several complexes of redox proteins that are thought to be highly dynamic, including Cyt b_5-myoglobin,[222] Cyt c-adrenodoxin,[223] Cyt c-Pc,[224] and Cyt c-Cyt b_5.[225] In this study, similarly small CSPs were observed in all Pc-peptide complexes. Small CSP can be caused by a dynamic interaction or simply low affinity,

such as a small fraction of bound peptide. In the case of PoPc and DPc, the CSP could be extrapolated to 100% on the basis of the K_D, demonstrating that the CSP are indeed small for the fully bound Pc. For PhPc the CSP were even too small to derive a reliable K_D. To support the hypothesis that small overall CSP values correlate with dynamic interactions, we used PRE mapping. The observed PREs are scattered over the Pc surface, and both for PoPc and DPc they cannot be satisfied by a single orientation of the peptides. Furthermore, the N- and C-terminal TOAC containing peptides gave very similar PRE maps, which is not to be expected for peptides binding in well-defined orientations. Thus, qualitatively the PRE results strongly support a dynamic binding model, in which the peptide assumes many orientations relative to Pc and interconverts between these orientations faster than the NMR timescale defined by the maximum CSP (exchange rate $>> 250$ s^{-1}).

Back-calculated distances using the ensemble docking approach with multiple orientations showed a good correlation with the experimental PREs for ensemble sizes much larger than 1, which is in line with dynamics within the complex. Also the average distances between TOAC and Pc amides of the MC dock ensembles matched the experimental distances qualitatively, but not quantitatively; the TOAC molecules were on average not close enough to the affected Pc amide groups to explain the observed PREs. This observation could be a consequence of the limitations of the docking method, such as the use of an exclusion grid to avoid steric hindrance. Alternatively, it could indicate small contributions of interactions other than electrostatics, perhaps very transient hydrogen bond formation between the exposed amide protons and the oxygen of TOAC. Evidence for the latter explanation comes from the PRE pattern. It is remarkable that the NMR resonances of several residues are broadened out beyond detection due to a PRE, whereas those of neighboring amides are (almost) unaffected. The distance between neighboring amides is about 4 Å, so the PRE ratio for two amide residues is at most proportional to $r^{-6}/(r+4)^{-6}$, where r is the distance between the nitroxy radical and the nearest amide proton. It can be shown that at least for some amides this must imply that the TOAC nitroxy group gets very close, within several Ångström for a short fraction of the time, which suggests that the sensitivity of PRE for minor states provides evidence for weak and transient short-range interactions. In physiological systems of protein-protein complexes such interactions must occur in the encounter complex next to the dominant charge-charge interactions for the complex to proceed to the final, well-defined complex.

5.5 Conclusions

The binding of tetralysine peptides to Pcs with different surface charge properties has been characterized by a combination of CSP, PRE NMR and MC simulations. The high similarity of CSP maps for the different peptides used in the study, as well as the small magnitudes of CSPs strongly suggests a high degree of dynamics. Also the scattered distribution of PREs indicates the presence of multiple orientations. The peculiar distribution of peptide positions obtained from ensemble docking with high densities in small areas only qualitatively matches the electrostatic docking simulations, suggesting that the PRE approach picks up very transient, short-range interactions between the peptide and the protein, in which the TOAC approaches specific amide protons very closely.

6 Concluding remarks

NMR in protein-ligand interactions

NMR is a powerful tool at various steps of drug discovery and development. From ligand screening, hit validation to structural characterization, NMR is commonly used to characterize the structure and dynamics of the protein target and the ligand.

Throughout this work, NMR spectroscopy has been extensively used to study molecular interactions. 1D ^{1}H-NMR was used in Chapters 2 and 4 in various ways, such as TINS (target-immobilized NMR screening) fragment screening, characterizing ligand binding with a T_2 relaxation filter, as well as paramagnetic pseudocontact shifts (PCSs). Isotope filtered/edited NOESY type experiments were used in Chapter 3. Chemical shift perturbation (CSP) analysis using [^{1}H,^{15}N]-HSQC titrations were used in Chapters 3 and 5 to determine binding constants of protein-small molecule and protein-peptide complexes. In addition, paramagnetic NMR spectroscopy was used to obtain PCS restraints in Chapter 4 (1D ^{1}H-NMR for ligand PCSs and [^{1}H,^{15}N]-HSQC for protein PCSs) and paramagnetic relaxation enhancement (PRE) restraints in Chapter 5 ([^{1}H,^{15}N]-HSQC). This demonstrated the robustness of classical NMR approaches and the never-ending evolution of new NMR techniques that complement the conventional NMR approaches in structural studies.

Similarity and difference between CPMG and PRE: relaxation, and relaxation

A T_2 relaxation filter can be used to filter out the sharp NMR signals of ligands from those of macromolecules, such as proteins. The filter can be applied using a Carr-Purcell-Meiboom-Gill (CPMG) or a spin-lock pulse sequence. When the ligand is bound to the protein, it becomes slow-tumbling, and consequently the signal intensity of the ligand is reduced. With a reasonable fraction of bound ligand, changes in chemical shifts induced by ligand binding can also be observed, if no extensive line-broadening is used in spectra processing. By observing the differences in chemical shifts as well as the intensities between the free and the bound state, the binding ligands can be distinguished from the non-binding ones. It

should be noted that, in contrast to PRE (discussed below), the broadening effect is global on the bound ligand, i.e. there is no distance dependence between the bound ligand nuclei and the protein nuclei. For molecules with labile protons, the exchange process can also have significant broadening effects on the labile protons, but not on the non-labile ones.

PRE NMR spectroscopy has proved to be a powerful technique for studying transient complexes.[120,121] PRE arises from the unpaired electrons, which have very strong magnetic moments. These unpaired electrons may enhance relaxation rates of nearby nuclei because of the dipolar interaction between the unpaired electron and the nucleus as well as the fast longitudinal relaxation of the electron spin. The paramagnetic center can be an intrinsic metal binding site, or an engineered external tag on the protein surface. For synthetic peptides, the paramagnetic amino acid TOAC can be incorporated in peptides via a peptide bond, making it very convenient in automated peptide synthesis. Like for the NOE, the magnitude of PRE is strongly dependent on the distance between the paramagnetic center and the observed nucleus (r^{-6}). Even lowly populated states which only exist shortly in the vicinity of the spin label can be observed. Therefore, similar to NOE in a certain extent, PRE is very sensitive to changes in distance. In contrast to NOE (effective range ~5 Å), PRE detects long range (10-35 Å) interaction with high sensitivity. Compared with the T_2 relaxation filter which simply removes broad signals based on the molecular size, PRE induces resonance broadening depending on the distance.

Comparison of protein-ligand structures determined by NOE and by PCS

The experimental data from NOESY experiments can be used as proton-proton distances between protein-ligand (intermolecular NOE) and within the protein residues (intramolecular NOE). With the use of filtered NOESY experiments, one can be used to distinguish protein-protein and protein-ligand NOE. In Chapter 3, this strategy was successfully applied to determine the ligand binding mode using 43 intermolecular NOE restraints. Given the large number of restraints with a nature of short-range interaction, the ligand binding mode can be determined with high resolution. Although robust, however, the requirement of nearly complete backbone and side chain assignments for the bound protein is time-consuming and sometimes error prone. Besides, $^{13}C/^{15}N$-labeled proteins must be produced in relatively large quantity compared with the small quantity used in other NMR techniques. This can be costly if many ligand bound complexes need to be studied.

Pseudocontact shifts (PCSs) are one of the most well-known paramagnetic effects. PCSs are distance and orientation dependent, and the effect is predictable. Significant PCSs can be

measured up to 60 Å away from the paramagnetic center with strong lanthanides, thus being an important source of long-range restraints, particularly for large proteins or complexes. In Chapter 4, we demonstrated that it is possible to determine the ligand binding site and obtain a low-resolution structure using only restraints from simple 1D-^{1}H NMR spectra. This requires attachment of two-armed paramagnetic lanthanide tags (CLaNP-5) on three positions of the protein via disulfide bond linkage, one tag at a time. Besides, the ligand has to be in fast exchange on the NMR time scale. The ligand position is determined relative to the different paramagnetic tensor frames. Lanthanide positions can be modeled and tensor magnitudes (axial and rhombic components) estimated with sufficient accuracy if the protein structure is available. In Chapter 4, a low-resolution structure of the ligand binding mode has been determined using 21 ligand PCSs from three sets of para- and diamagnetic 1D-^{1}H NMR spectra. The final structure is similar to the structure derived from NOE restraints (Chapter 3). The quality of the structure can be improved if the tensor positions are optimized using experimental protein PCSs. Nevertheless, the structure information obtained with the predicted tensors is sufficient for the early stage of drug discovery. The advantage of this approach is that protein isotope labeling and protein backbone assignments are not compulsory. Most important are the ligand assignments, which are typically much easier to obtain. The resolution can further be improved by increasing the number of restraints. This can be achieved by using more than three tagging sites, using a tag with different tensor frames (such as CLaNP-7[188]), or measuring additional PCSs of the bound ligand in ^{13}C-, ^{19}F-, or ^{31}P-NMR spectra (if the ligand has fluorine or phosphorous atoms). Sometimes introducing two additional cysteines on the protein can affect the stability and structure of the protein. Therefore, it is advised to prepare more mutants than actually needed, and select the ones which are least affected by the extra cysteines and by the lanthanide tags.

Chemical shift perturbation analysis in protein-ligand complexes

CSP analysis has been frequently used to study protein-protein and protein-ligand interactions. The interaction is accompanied by changes in the chemical environment of the observed nuclei at the binding interface as well as at a remote region. This information is subsequently converted into binding affinities (K_D), and used to qualitatively estimate the degree of dynamics. CSPs are averaged over all orientations. The absolute size of CSPs has also been reported to correlate with the degree of dynamics: smaller chemical shift changes indicates higher mobility.[165,222,225,226] Observation of distant conformational changes indirectly caused by ligand binding has been shown in the example of FKBP12 bound to a

small molecule (Chapter 3). In Chapter 5, the overall small size of chemical shift perturbations strongly suggests that the Pc-tetralysine peptide complexes are highly dynamic. Therefore CSP analysis can be used as a tool to study protein-small molecule and protein-peptide dynamics. Further developments in the quantitative analysis of CSP may be needed to draw a clear line between the encounter complex and the specific complex.

Dynamics in transient protein-peptide complexes

To study interactions in transient complexes, the traditional NOE method is difficult to apply. The sensitivity of PRE to lowly populated states makes it a versatile technique to study dynamics in complex formation.[120] On the spectroscopy side, the interactions can be detected by simply acquiring standard HSQC spectra. On the sample preparation side, if the molecule (a protein or a ligand) is diamagnetic, an additional attachment step is required to introduce the paramagnetic source. It is also possible to replace the metal ion in metalloproteins with a suitable ion, to suppress or introduce paramagnetism to the system. In chapter 5, the electron transfer protein, plastocyanin (Pc), naturally harbours to a copper ion. In our *in vitro* study, the copper was replaced by a zinc ion to suppress PRE caused by the copper (II). The paramagnetic amino acid, TOAC, was attached on the peptide as the source of PRE. PRE data represent all orientations of the partners in the complex, including the sparse states. For encounter complexes that are mainly driven by long range electrostatic interactions, multiple orientations can exist with the same energy, like 'velcro'.[197] By comparing CSP and PRE maps of the Pc-tetralysine peptide complexes with the results from Monte Carlo (MC) simulations, it was found that the results of CSP maps and those of MC simulations were in high agreement for Pc from *Populus nigra* and *Dryopteris crassirhizoma*, which have many negative charges on the protein surface. For complexes with weaker charge complementarity, such as tetralysine binding to Pc from *Phormidium laminosum*, the binding is weak as observed in CSP and PRE data. The PRE data provide indisputable evidence that the peptide binds in multiple orientations.

Computational tools for studying molecular interactions

Modeling the positions of the lanthanide binding tag CLaNP-5 was first introduced in pseudoazurin and showed high accuracy.[144] The modeling approach is simple and can be valuable for proteins that cannot be produced with isotope labels. If the secondary structure of the tagging site is flexible, fluctuations due to the mobility of the residues should be considered. In Chapter 4, the binding mode of the small-molecule ligand bound to FKBP12

was calculated relative to the positions of the lanthanides, which were derived from modeling and later optimized using experimental data.

PCSdock (Chapter 4) provides a fast, rough estimate of the ligand binding site based on ligand PCSs. This simple procedure uses ligand PCSs as the input restraints, provided the tensor parameters and the protein structure are available. The ligand structure and assignments are not required. PCSdock simply shows the grid points that satisfy the input PCSs, which leads to a grid in which the ligand is likely to bind.

An ensemble docking approach (Chapter 5) was used to visualize the PRE data of the complexes of Pc-charged peptides. For specific complexes, PRE should represent a single orientation of the binding partners. For encounter complexes that can exist in multiple orientations, interpretation of PRE data using an ensemble of orientations provides better visualization than using a single orientation. However, the ensemble docking approach still provides limited solutions. As MC simulations consider only electrostatic interactions, the results of experimental PRE of complexes that are purely driven by electrostatics are expected to correlate well with those of MC simulations. In our study, the results of theoretical MC simulations with experimental CSP and PRE analysis on the charged peptide and the three Pcs with different surface charge properties provided evidence for dominant non-specific electrostatic interactions as well as other transient short-range interactions, such as hydrogen bonds and hydrophobic contacts.

References

(1) Siekierka, J. J.; Hung, S. H. Y.; Poe, M.; Lin, C. S.; Sigal, N. H. *Nature* **1989**, *341*, 755−757.

(2) Harding, M. W.; Galat, A.; Uehling, D. E.; Schreiber, S. L. *Nature* **1989**, *341*, 758–760.

(3) Marx, S. O.; Jayaraman, T.; Go, L. O.; Marks, A. R. *Circ. Res.* **1995**, *76*, 412–417.

(4) Liu, J.; Farmer, J. D.; Lane, W. S.; Friedman, J.; Weissman, I.; Schreiber, S. L. *Cell* **1991**, *66*, 807–815.

(5) Bierer, B. E.; Somers, P. K.; Wandless, T. J.; Burakoff, S. J.; Schreiber, S. L. *Science* **1990**, *250*, 556–559.

(6) Bierer, B. E.; Mattila, P. S.; Standaert, R. F.; Herzenberg, L. A.; Burakoff, S. J.; Crabtree, G. R.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9231–9235.

(7) Rosen, M. K.; Michnick, S. W.; Karplus, M.; Schreiber, S. L. *Biochemistry* **1991**, *30*, 4774–4789.

(8) Moore, J. M.; Peattie, D. A.; Fitzgibbon, M. J.; Thomson, J. A. *Nature* **1991**, *351*, 248–250.

(9) Michnick, S. W.; Rosen, M. K.; Wandless, T. J.; Karplus, M.; Schreiber, S. L. *Science* **1991**, *252*, 836–839.

(10) Petros, A. M.; Kawai, M.; Luly, J. R.; Fesik, S. W. *FEBS Lett.* **1992**, *308*, 309–314.

(11) Xu, R. X.; Olejniczak, E. T.; Fesik, S. W. *FEBS Lett.* **1992**, *305*, 137–143.

(12) Xu, R. X.; Nettesheim, D.; Meadows, R.; Gemmecker, G.; Fesik, S. W. *Biopolymers* **1993**, *33*, 535−550.

(13) Meadows, R. P.; Nettesheim, D. G.; Xu, R. X.; Olejniczak, E. T.; Petros, A. M.; Holzman, T. F.; Severin, J.; Gubbins, E.; Smith, H. T.; Fesik, S. W. *Biochemistry* **1993**, *32*, 754–765.

(14) Sich, C.; Improta, S.; Cowley, D. J.; Guenet, C.; Merly, J.-P.; Teufel, M.; Saudek, V. *Eur. J. Biochem.* **2000**, *267*, 5342–5355.

(15) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *Science* **1991**, *179*, 839−842.

(16) Van Duyne, G. D.; Standaert, R. F.; Schreiber, S. L.; Clardy, J. *J. Am. Chem. Soc.* **1991**, *113*, 7434−7435.

(17) Sayal, G.; Vplkin, D. B.; Marcy, A. I.; Chan, H. K.; Ryan, J. A.; Middaugh, C. R. *Biochem. Biophys. Res. Commun.* **1991**, *179*, 741–748.

(18) Holt, D. A.; Luengo, J. J. I.; Yamashita, D. S.; Oh, H.; Konialian, A. L.; Yen, H.; Rozamus, L. W.; Brandt, M.; Bossard, M. J.; Levy, M. A.; Eggleston, D. S.; Liang, J.; Schultz, L. W.; Stout, T. J.; Clardy, J. *J. Am. Chem. Soc.* **1993**, 9925–9938.

(19) Becker, J. W.; Rotonda, J.; Mckeevers, B. M.; Chan, H. K.; Marcy, A. I.; Wiederrecht, G.; Hermess, J. D.; Springer, J. P. *J. Biol. Chem.* **1993**, *268*, 11335–11339.

(20) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *J. Mol. Biol.* **1993**, *229*, 105–124.

(21) Itoh, S.; DeCenzo, M. T.; Livingston, D. J.; Perlman, D. A.; Navia, M. A. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 1983–1988.

(22) Griffith, J. P.; Kim, J. L.; Kim, E. E.; Sintchak, M. D.; Thomson, J. A.; Fitzgibbon, M. J.; Fleming, M. A.; Caron, P. R.; Hsiao, K.; Navia, M. A. *Cell* **1995**, *82*, 507–522.

(23) Choi, J.; Chen, J.; Schreiber, S. L.; Clardy, J. *Science* **1996**, *273*, 239−242.

(24) Schultz, L. W.; Clardy, J. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1−6.

(25) Liang, J.; Choi, J.; Clardy, J. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **1999**, *55*, 736−744.

(26) Huse, M.; Chen, Y. *Cell* **1999**, *96*, 425–436.

(27) Burkhard, P.; Taylor, P.; Walkinshaw, M. D. *J. Mol. Biol.* **2000**, *295*, 953–962.

(28) Sun, F.; Li, P.; Ding, Y.; Wang, L.; Bartlam, M.; Shu, C.; Shen, B.; Jiang, H.; Li, S.; Rao, Z. *Biophys. J.* **2003**, *85*, 3194–3201.

(29) Szep, S.; Park, S.; Boder, E. T.; van Duyne, G. D.; Saven, J. G. *Proteins* **2009**, *74*, 603–611.

(30) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531–1534.

(31) Barber, J. *Plant Cell Environ.* **1983**, *6*, 311–322.

(32) Haehnel, W. *Ann. Rev. Plant Physiol.* **1984**, *35*, 659–693.

(33) Barber, J.; Anderson, B. *Nature* **1994**, *370*, 31–34.

(34) Nugent, J. H. *Eur. J. Biochem.* **1996**, *237*, 519–531.

(35) Chapman, G.; Colman, P. M.; Freeman, H. C.; Guss, J. M.; Murata, M.; Norris, V. A.; Ramshaw, J. A. M.; Venkatappa, M. P. *J. Mol. Biol.* **1977**, *110*, 187–189.

(36) Guss, J. M.; Harrowell, P. R.; Murata, M.; Norris, V. A.; Freeman, H. C. *J. Mol. Biol.* **1986**, *192*, 361–387.

(37) Moore, J. M.; Case, D. A.; Chazin, W. J.; Gippert, G. P.; Havel, T. F.; Powls, R.; Wright, P. E. *Science* **1988**, *240*, 314–317.

(38) Driscoll, P. C.; Hill, H. A.; Redfield, C. *Eur. J. Biochem.* **1987**, *292*, 279−292.

(39) King, G. C.; Wright, P. E. *Biochemistry* **1986**, *25*, 2364–2374.

(40) Chazin, W. J.; Wright, P. E. *J. Mol. Biol.* **1988**, *202*, 623–636.

(41) Moore, J. M.; Lepre, C. A.; Gippert, G. P.; Chazin, W. J.; Case, D. A.; Wright, P. E. *J. Mol. Biol.* **1991**, *221*, 533–555.

(42) Bond, C. S.; Bendall, D. S.; Freeman, H. C.; Guss, J. M.; Howe, C. J.; Wagner, M. J.; Wilce, M. C. J. *Acta Cryst.* **1999**, *D55*, 414–421.

(43) Kohzuma, T.; Inoue, T.; Yoshizaki, F.; Sasakawa, Y.; Onodera, K.; Nagatomo, S.; Kitagawa, T.; Uzawa, S.; Isobe, Y.; Sugimura, Y.; Gotowda, M.; Kai, Y. *J. Biol. Chem.* **1999**, *274*, 11817−11823.

(44) Hirota, S.; Hayamizu, K.; Okuno, T.; Kishi, M.; Iwasaki, H.; Kondo, T. *Biochemistry* **2000**, *39*, 6357–6364.

(45) Garrett, T. P. J.; Clingeleffer, D. J.; Guss, J. M.; Rogers, S. J.; Freeman, H. C. *J. Biol. Chem.* **2006**, *259*, 2822−2825.

(46) Hulsker, R.; Mery, A.; Thomassen, E. A.; Ranieri, A.; Sola, M.; Verbeet, M. P.; Kohzuma, T.; Ubbink, M. *J. Am. Chem. Soc.* **2007**, *129*, 4423–4429.

(47) Lange, C.; Cornvik, T.; Díaz-Moreno, I.; Ubbink, M. *Biochim. Biophys. Acta* **2005**, *1707*, 179–188.

References

(48) Crowley, P. B.; Hunter, D. M.; Sato, K.; McFarlane, W.; Dennison, C. *Biochem. J.* **2004**, *378*, 45–51.

(49) Jansson, H.; Ökvist, M.; Jacobson, F.; Ejdebäck, M.; Hansson, Ö.; Sjölin, L. *Biochim. Biophys. Acta* **2003**, *1607*, 203−210.

(50) Guss, J. M.; Bartunik, H. D.; Freeman, H. C. *Acta Cryst.* **1992**, *B48*, 790−811.

(51) Colman, P. M.; Freeman, H. C.; Guss, J. M.; Murata, M.; Norris, V. A.; Ramshaw, J. A. M.; Venkatappa, M. P. *Nature* **1978**, *272*, 319–324.

(52) Guss, J. M.; Freeman, H. C. *J. Mol. Biol.* **1983**, *169*, 521–563.

(53) Bagby, S.; Driscoll, P. C.; Harvey, T. S.; Hill, H. A. O. *Biochemistry* **1994**, *33*, 6611–6622.

(54) Xue, Y.; Ökvist, M.; Hansson, Ö.; Young, S. *Protein Sci.* **1998**, *7*, 2099–2105.

(55) Congreve, M.; Carr, R. A. E.; Murray, C.; Jhoti, H. *Drug Discov. Today* **2003**, *8*, 876–877.

(56) Hopkins, A. L.; Groom, C. R. *Drug Discov. Today* **2004**, *9*, 430–431.

(57) Hubbard, R. E.; Murray, J. B. *Methods Enzymol.* **2011**, *493*, 509−531.

(58) Winter, A.; Higueruelo, A. P.; Marsh, M.; Sigurdardottir, A.; Pitt, W. R.; Blundell, T. L. *Q. Rev. Biophys.* **2012**, *45*, 383–426.

(59) Scott, D. E.; Coyne, A. G.; Hudson, S. a; Abell, C. *Biochemistry* **2012**, *51*, 4990–5003.

(60) Blundell, T. L.; Jhoti, H.; Abell, C. *Nat. Rev. Drug Discov.* **2002**, *1*, 45−54.

(61) De Mol, N. J.; Fischer, M. J. E. *Methods Mol. Biol.* **2010**, *627*, 1–14.

(62) Cooper, M. A. *Nat. Rev. Drug Discov.* **2002**, *1*, 515–528.

(63) Freire, E. *Drug Discov. Today* **2008**, *13*, 869–874.

(64) Ladbury, J. E.; Klebe, G.; Freire, E. *Nat. Rev. Drug Discov.* **2010**, *9*, 23–27.

(65) Silvestre, H. L.; Blundell, T. L.; Abell, C.; Ciulli, A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 12984–12989.

(66) Pellecchia, M.; Sem, D. S.; Wüthrich, K. *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219.

(67) Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. *Nat. Rev. Drug Discov.* **2008**, *7*, 738–745.

(68) Stockman, B. J.; Dalvit, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41*, 187–231.

(69) Homans, S. W. *Angew. Chem. Int. Ed.* **2004**, *43*, 290−300.

(70) Diercks, T.; Coles, M.; Kessler, H. *Curr. Opin. Chem. Biol.* **2001**, *5*, 285–291.

(71) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366–12371.

(72) Wider, G. *Methods Enzymol.* **2005**, *394*, 382–398.

(73) Hajduk, P. J.; Augeri, D. J.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S. F.; Fesik, S. W.; Park, A. *J. Am. Chem. Soc.* **2000**, *122*, 7898–7904.

(74) Tugarinov, V.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 13868–13878.

(75) Hajduk, P. J.; Burns, D. J. *Comb. Chem. High Throughput Screen.* **2002**, *5*, 613–621.

(76) Mayer, M.; Meyer, B. *Angew. Chem. Int. Ed.* **1999**, *35*, 1784–1788.

(77) Mayer, M.; Meyer, B. *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.

(78) Jayalakshmi, V.; Krishna, N. R. *J. Am. Chem. Soc.* **2005**, *127*, 14080–14084.

(79) Jayalakshmi, V.; Biet, T.; Peters, T.; Krishna, N. R. *J. Am. Chem. Soc.* **2004**, *126*, 8610–8611.

(80) Hajduk, P. J.; Mack, J. C.; Olejniczak, E. T.; Park, C.; Dandliker, P. J.; Beutel, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 2390−2398.

(81) Dalvit, C.; Pevarello, P.; Tatò, M.; Veronesi, M.; Vulpetti, A.; Sundström, M. *J. Biomol. NMR* **2000**, *18*, 65–68.

(82) Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. J. *J. Biomol. NMR* **2001**, *21*, 349–359.

(83) Dalvit, C.; Fasolini, M.; Flocco, M.; Knapp, S.; Pevarello, P.; Veronesi, M. *J. Med. Chem.* **2002**, *45*, 2610−2614.

(84) Vanwetswinkel, S.; Heetbrij, R. J.; Duynhoven, J. van; Hollander, J. G.; Filippov, D. V; Hajduk, P. J.; Siegal, G. *Chem. Biol.* **2005**, *12*, 207–216.

(85) Früh, V.; Zhou, Y.; Chen, D.; Loch, C.; AB, E.; Grinkova, Y. N.; Verheij, H.; Sligar, S. G.; Bushweller, J. H.; Siegal, G. *Chem. Biol.* **2010**, *17*, 881–889.

(86) Congreve, M.; Rich, R. L.; Myszka, D. G.; Figaroa, F.; Siegal, G.; Marshall, F. H. *Methods Enzymol.* **2011**, *493*, 115–136.

(87) Chen, D.; Errey, J. C.; Heitman, L. H.; Marshall, F. H.; Ijzerman, A. P.; Siegal, G. *ACS Chem. Biol.* **2012**, *7*, 2064–2073.

(88) Sánchez-Pedregal, V. M.; Reese, M.; Meiler, J.; Blommers, M. J. J.; Griesinger, C.; Carlomagno, T. *Angew. Chem. Int. Ed.* **2005**, *44*, 4172–4175.

(89) Keizers, P. H. J.; Ubbink, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 88–96.

(90) Otting, G. *Annu. Rev. Biophys.* **2010**, *39*, 387−405.

(91) Otting, G. *J. Biomol. NMR* **2008**, *42*, 1–9.

(92) Jahnke, W.; Rüdisser, S.; Zurini, M. *J. Am. Chem. Soc.* **2001**, *123*, 3149–3150.

(93) Jahnke, W.; Perez, L. B.; Paris, C. G.; Strauss, A.; Fendrich, G.; Nalin, C. M. *J. Am. Chem. Soc.* **2000**, *122*, 7394–7395.

(94) Jahnke, W.; Blommers, M. J. J.; Fernández, C.; Zwingelstein, C.; Amstutz, R. *ChemBioChem* **2005**, *6*, 1607–1610.

(95) Bertini, I.; Fragai, M.; Lee, Y.-M.; Luchinat, C.; Terni, B. *Angew. Chem. Int. Ed.* **2004**, *43*, 2254–2256.

(96) Zhuang, T.; Lee, H.-S.; Imperiali, B.; Prestegard, J. H. *Protein Sci.* **2008**, *17*, 1220–1231.

(97) John, M.; Pintacuda, G.; Park, A. Y.; Dixon, N. E.; Otting, G. *J. Am. Chem. Soc.* **2006**, *128*, 12910−12916.

(98) Gochin, M.; Zhou, G.; Phillips, A. H. *ACS Chem. Biol.* **2011**, *6*, 267–274.

(99) Saio, T.; Ogura, K.; Shimizu, K.; Yokochi, M.; Burke, T. R.; Inagaki, F. *J. Biomol. NMR* **2011**, *51*, 395–408.

(100) van Dijk, A. D. J.; Boelens, R.; Bonvin, A. M. J. J. *FEBS J.* **2005**, *272*, 293–312.

(101) Hunter, C. A.; Packer, M. J.; Zonta, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *47*, 27–39.

References

(102) Williamson, M. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.

(103) González-Ruiz, D.; Gohlke, H. *J. Chem. Inf. Model.* **2009**, *49*, 2260–2271.

(104) Cioffi, M.; Hunter, C. A.; Packer, M. J.; Pandya, M. J.; Williamson, M. P. *J. Biomol. NMR* **2009**, *43*, 11–19.

(105) Cioffi, M.; Hunter, C. A.; Packer, M. J.; Spitaleri, A. *J. Med. Chem.* **2008**, *51*, 2512–2517.

(106) Cioffi, M.; Hunter, C. A.; Packer, M. J. *J. Med. Chem.* **2008**, *51*, 4488–4495.

(107) Stark, J.; Powers, R. *J. Am. Chem. Soc.* **2008**, *130*, 535–545.

(108) Constantine, K. L.; Davis, M. E.; Metzler, W. J.; Mueller, L.; Claus, B. L. *J. Am. Chem. Soc.* **2006**, *128*, 7252–7263.

(109) Su, X.-C.; Otting, G. *J. Biomol. NMR* **2010**, *46*, 101–112.

(110) Rodriguez-Castañeda, F.; Haberz, P.; Leonov, A.; Griesinger, C. *Magn. Reson. Chem.* **2006**, *44*, S10−S16.

(111) Koehler, J.; Meiler, J. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 360–389.

(112) Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009**, *109*, 4108−4139.

(113) Bertini, I.; Luchinat, C.; Parigi, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249–273.

(114) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *ChemBioChem* **2005**, *6*, 1536–1549.

(115) Bertini, I.; Luchinat, C.; Piccioli, M. *Methods Enzymol.* **2001**, *339*, 314–340.

(116) Bertini, I.; Turano, P.; Vila, A. J. *Chem. Rev.* **1993**, *93*, 2833–2932.

(117) Keizers, P. H. J.; Ubbink, M. In *Protein NMR Spectroscopy: Practical Techniques and Applications*; Lian, L.-Y.; Roberts, G., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2011; pp. 193–219.

(118) Bertini, I.; Luchinat, C.; Parigi, G. *Solution NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models*; 1st ed.; Elsevier Science B.V.: Amsterdam, 2001.

(119) Bertini, I.; Luchinat, C.; Parigi, G. In *NMR of Biomolecules: Towards Mechanistic System Biology*; Bertini, I.; McGreevy, K. S.; Parigi, G., Eds.; Wileu-VCH Verlag GmbH & Co. KGaA., 2012; pp. 155–171.

(120) Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006**, *444*, 383–386.

(121) Iwahara, J.; Clore, G. M. *Nature* **2006**, *440*, 1227−1230.

(122) Grimaldi, J. J.; Grant, T. *J. Biol. Chem.* **1975**, *250*, 1618–1624.

(123) Lee, L.; Sykes, B. D.; Birnbaum, E. R. *FEBS Lett.* **1979**, *98*, 169–172.

(124) Bentrop, D.; Bertini, I.; Cremonini, M. A.; Forsén, S.; Luchinat, C.; Malmendal, A. *Biochemistry* **1997**, *36*, 11605–11618.

(125) Bertini, I.; Gupta, Y. K.; Luchinat, C.; Parigi, G.; Peana, M.; Sgheri, L.; Yuan, J. *J. Am. Chem. Soc.* **2007**, *129*, 12786–12794.

(126) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *Dalt. Trans.* **2008**, 3782–3790.

(127) Bertini, I.; Kursula, P.; Luchinat, C.; Parigi, G.; Vahokoski, J.; Wilmanns, M.; Yuan, J. *J. Am. Chem. Soc.* **2009**, *131*, 5134–5144.

(128) Beattie, J. K.; Fensom, D. J.; Freeman, H. C.; Woodcock, E.; Hill, H. A. O.; Stokes, A. M. *Biochim. Biophys. Acta* **1975**, *405*, 109–114.

(129) Shulman, R. G.; Wüthrich, K.; Yamane, T.; Antonini, E.; Brunori, M. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *63*, 623–628.

(130) Franz, K. J.; Nitz, M.; Imperiali, B. *ChemBioChem* **2003**, *4*, 265–271.

(131) Clark, I. D.; Hill, I.; Sikorska-Walker, M.; MacManus, J. P.; Szabo, A. G. *FEBS Lett.* **1993**, *333*, 96–98.

(132) Nitz, M.; Sherawat, M.; Franz, K. J.; Peisach, E.; Allen, K. N.; Imperiali, B. *Angew. Chem. Int. Ed.* **2004**, *43*, 3682–3685.

(133) Su, X.-C.; McAndrew, K.; Huber, T.; Otting, G. *J. Am. Chem. Soc.* **2008**, *130*, 1681−1687.

(134) Saio, T.; Ogura, K.; Yokochi, M.; Kobashigawa, Y.; Inagaki, F. *J. Biomol. NMR* **2009**, *44*, 157−166.

(135) Barthelmes, K.; Reynolds, A. M.; Peisach, E.; Jonker, H. R. a; DeNunzio, N. J.; Allen, K. N.; Imperiali, B.; Schwalbe, H. *J. Am. Chem. Soc.* **2011**, *133*, 808–819.

(136) Su, X.-C.; Huber, T.; Dixon, N. E.; Otting, G. *ChemBioChem* **2006**, *7*, 1599–1604.

(137) Xie, J.; Liu, W.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2007**, *46*, 9239−9242.

(138) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225−249.

(139) Jones, D. H.; Cellitti, S. E.; Hao, X.; Zhang, Q.; Jahnz, M.; Summerer, D.; Schultz, P. G.; Uno, T.; Geierstanger, B. H. *J. Biomol. NMR* **2010**, *46*, 89−100.

(140) Griffith, O. H.; McConnel, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 8–11.

(141) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923−2938.

(142) Bettio, A.; Pöppl, V.; Andreas, G.; Dinger, M. C.; Zschörnig, O.; Arnold, K.; Toniolo, C.; Beck-Sickinger, A. G. *J. Pept. Sci.* **2002**, *8*, 671–682.

(143) Lindfors, H. E.; de Koning, P. E.; Drijfhout, J. W.; Venezia, B.; Ubbink, M. *J. Biomol. NMR* **2008**, *41*, 157–167.

(144) Keizers, P. H. J.; Saragliadis, A.; Hiruma, Y.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2008**, *130*, 14802–14812.

(145) Hirota, S.; Okumura, H.; Arie, S.; Tanaka, K.; Shionoya, M.; Takabe, T.; Funasaki, N.; Watanabe, Y. *J. Inorg. Biochem.* **2004**, *98*, 849−855.

(146) Hirota, S.; Yamauchi, O. *Eur. J. Inorg. Chem.* **2002**, 17−25.

(147) Hirota, S.; Hayamizu, K.; Endo, M.; Hibino, T.; Takabe, T.; Kohzuma, T.; Yamauchi, O. *J. Am. Chem. Soc.* **1998**, *120*, 8177−8183.

(148) Lepre, C. A. *Methods Enzymol.* **2011**, *493*, 219–239.

(149) Sun, C.; Petros, A. M.; Hajduk, P. J. *J. Comput. Mol. Des.* **2011**, *25*, 607–610.

(150) Powers, R. *J. Struct. Funct. Genomics* **2002**, *2*, 113–123.

(151) Röhrig, C. H.; Loch, C.; Guan, J.-Y.; Siegal, G.; Overhand, M. *ChemMedChem* **2007**, *2*, 1054–1070.

(152) Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. *J. Med. Chem.* **1999**, *42*, 2498−2503.

(153) Schüttelkopf, A. W.; van Aalten, D. M. F. *Acta Cryst.* **2004**, *D60*, 1355–1363.

(154) Aspinall, H. C. *Chem. Rev.* **2002**, *102*, 1807–1850.

(155) Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783–2792.

(156) Nilsson, M.; Morris, G. A. *Chem. Commun.* **2007**, 933−935.

(157) Aguilar, J. A.; Faulkner, S.; Nilsson, M.; Morris, G. A. *Angew. Chem. Int. Ed.* **2010**, *49*, 3901−3903.

(158) Aguilar, J. A.; Nilsson, M.; Morris, G. A. *Angew. Chem. Int. Ed.* **2011**, *50*, 9716−9717.

(159) Otting, G.; Wüthrich, K. *Q. Rev. Biophys.* **1990**, *23*, 39–96.

(160) Stuart, A. C.; Borzilleri, K. A.; Withka, J. M.; Palmer, A. G. *J. Am. Chem. Soc.* **1999**, *121*, 5346–5347.

(161) Breeze, A. L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2000**, *36*, 323–372.

(162) Wandless, T. J.; Michnick, S. W.; Rosen, M. K.; Karplus, M.; Schreiber, S. L. *J. Am. Chem. Soc.* **1991**, *113*, 2339−2341.

(163) Brath, U.; Akke, M. *J. Mol. Biol.* **2009**, *387*, 233–244.

(164) Kneller, T. D.; Goddard, D. G. SPARKY 3.

(165) Worrall, J. A. R.; Reinle, W.; Bernhardt, R.; Ubbink, M. *Biochemistry* **2003**, *450*, 7068–7076.

(166) Kannt, A.; Young, S.; Bendall, D. S. *Biochim. Biophys. Acta* **1996**, *1277*, 115–126.

(167) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283–298.

(168) Sapienza, P. J.; Mauldin, R. V; Lee, A. L. *J. Mol. Biol.* **2011**, *405*, 378–394.

(169) Schrödinger, L. The PyMOL Molecular Graphics System, Version 0.99.

(170) Chuang, W.-T.; Hsieh, C.-C.; Lai, C.-H.; Lai, C.-H.; Shih, C.-W.; Chen, K.-Y.; Hung, W.-Y.; Hsu, Y.-H.; Chou, P.-T. *J. Org. Chem.* **2011**, *76*, 8189–8202.

(171) Cheng, J.-W.; Lepre, C. A.; Moore, J. M. *Biochemistry* **1994**, *33*, 4093–4100.

(172) Cheng, J.-W.; Lepre, C. A.; Chambers, S. P.; Fulghum, R.; Thomson, J. A.; Moore, J. M. *Biochemistry* **1993**, *32*, 9000–9010.

(173) Brath, U.; Akke, M.; Yang, D.; Kay, L. E.; Mulder, F. A. A. *J. Am. Chem. Soc.* **2006**, *128*, 5718–5727.

(174) Hajduk, P. J.; Greer, J. *Nat. Rev. Drug Discov.* **2007**, *6*, 211–219.

(175) Schieborr, U.; Vogtherr, M.; Elshorst, B.; Betz, M.; Grimme, S.; Pescatore, B.; Langer, T.; Saxena, K.; Schwalbe, H. *ChemBioChem* **2005**, *6*, 1891–1898.

(176) Keizers, P. H. J.; Desreux, J. F.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2007**, *129*, 2–5.

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

(178) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn.Res.* **2003**, *160*, 65–73.

(179) Banci, L.; Bertini, I.; Cavallaro, G.; Giachetti, A.; Luchinat, C.; Parigi, G. *J. Biomol. NMR* **2004**, *28*, 249–261.

(180) Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010**, *132*, 241–247.

(181) Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open Source Scientific Tools for Python **2001**.

(182) Wöhnert, J.; Franz, K. J.; Nitz, M.; Imperiali, B.; Schwalbe, H. *J. Am. Chem. Soc.* **2003**, *125*, 13338–13339.

(183) Martin, L. J.; Hähnke, M. J.; Nitz, M.; Wöhnert, J.; Silvaggi, N. R.; Allen, K. N.; Schwalbe, H.; Imperiali, B. *J. Am. Chem. Soc.* **2007**, *129*, 7106–7113.

(184) Xu, X.; Keizers, P. H. J.; Reinle, W.; Hannemann, F.; Bernhardt, R.; Ubbink, M. *J. Biomol. NMR* **2009**, *43*, 247–254.

(185) Hass, M. A. S.; Keizers, P. H. J.; Blok, A.; Hiruma, Y.; Ubbink, M. *J. Am. Chem. Soc.* **2010**, *132*, 9952−9953.

(186) Bertini, I.; Calderone, V.; Cerofolini, L.; Fragai, M.; Geraldes, C. F. G. C.; Hermann, P.; Luchinat, C.; Parigi, G.; Teixeira, J. M. C. *FEBS Lett.* **2012**, *586*, 557–567.

(187) Dasgupta, S.; Hu, X.; Keizers, P. H. J.; Liu, W.-M.; Luchinat, C.; Nagulapalli, M.; Overhand, M.; Parigi, G.; Sgheri, L.; Ubbink, M. *J. Biomol. NMR* **2011**, *51*, 253–263.

(188) Liu, W.-M.; Keizers, P. H. J.; Hass, M. A. S.; Blok, A.; Timmer, M.; Sarris, A. J. C.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2012**, *134*, 17306−17313.

(189) Longinetti, M.; Parigi, G.; Sgheri, L. *J. Phys. A: Math. Gen.* **2002**, *35*, 8153–8169.

(190) Bertini, I.; Longinetti, M.; Luchinat, C.; Parigi, G.; Sgheri, L. *J. Biomol. NMR* **2002**, *22*, 123–136.

(191) Saio, T.; Yokochi, M.; Kumeta, H.; Inagaki, F. *J. Biomol. NMR* **2010**, *46*, 271–280.

(192) Bertini, I.; Fragai, M.; Giachetti, A.; Luchinat, C.; Maletta, M.; Parigi, G.; Yeo, K. J. *J. Med. Chem.* **2005**, *48*, 7544–7559.

(193) Pavlopoulou, A.; Michalopoulos, I. *Int. J. Mol. Med.* **2011**, *28*, 295–310.

(194) Dhaliwal, B.; Chen, Y. W. *Infect. Disord. Drug Targets* **2009**, *9*, 557–562.

(195) Bashir, Q.; Scanu, S.; Ubbink, M. *FEBS J.* **2011**, *278*, 1391–1400.

(196) Ubbink, M. *FEBS Lett.* **2009**, *583*, 1060–1066.

(197) McLendon, G. *Struct. Bond.* **1991**, *75*, 159–174.

(198) Sugawara, H.; Inoue, T.; Li, C.; Gotowda, M.; Hibino, T.; Takabe, T.; Kai, Y. *J. Biochem.* **1999**, *125*, 899–903.

(199) Collyer, C. A.; Guss, J. M.; Sugimura, Y.; Yoshizaki, F.; Freeman, H. C. *J. Mol. Biol.* **1990**, *211*, 617–632.

(200) Redinbo, M. R.; Cascio, D.; Choukair, M. K.; Rice, D.; Merchant, S.; Yeates, T. O. *Biochemistry* **1993**, *32*, 10560–10567.

(201) Shibata, N.; Inoue, T.; Nagano, C.; Nishio, N.; Kohzuma, T.; Onodera, K.; Yoshizaki, F.; Sugimura, Y.; Kai, Y. *J. Biol. Chem.* **1999**, *274*, 4225–4230.

(202) Bertini, I.; Bryant, D. A.; Ciurli, S.; Dikiy, A.; Fernández, C. O.; Luchinat, C.; Safarov, N.; Vila, A. J.; Zhao, J. *J. Biol. Chem.* **2001**, *276*, 47217–47226.

(203) Schmidt, L.; Christensen, H. E. M.; Harris, P. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **2006**, *62*, 1022–1029.

(204) Díaz-Moreno, I.; Díaz-Quintana, A.; De la Rosa, M. A.; Ubbink, M. *J. Biol. Chem.* **2005**, *280*, 18908–18915.

(205) Hulsker, R.; Baranova, M. V; Bullerjahn, G. S.; Ubbink, M. *J. Am. Chem. Soc.* **2008**, *130*, 1985−1991.

(206) Hirota, S.; Endo, M.; Tsukazaki, T. *J. Biol. Inorg. Chem* **1998**, *3*, 563–569.

(207) Hirota, S.; Endo, M. *J. Am. Chem. Soc.* **1999**, *121*, 849–855.

(208) Hirota, S.; Tsukazaki, T.; Yamauchi, O. *Biochem. Biophys. Res. Commun.* **2000**, *268*, 395–397.

(209) Hirota, S.; Yamauchi, O. *Chem. Rec.* **2001**, *1*, 290−299.

(210) Schreier, S.; Bozelli, J. C.; Marín, N.; Vieira, R. F. F.; Nakaie, C. R. *Biophys. Rev.* **2012**, *4*, 45–66.

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

(212) Cai, M.; Huang, Y.; Sakaguchi, K.; Clore, G. M.; Gronenborn, A. M.; Craigie, R. *J. Biomol. NMR* **1998**, *11*, 97–102.

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

(214) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355–5365.

(215) Scanu, S.; Foerster, J. M.; Finiguerra, M. G.; Shabestari, M. H.; Huber, M.; Ubbink, M. *ChemBioChem* **2012**, *13*, 1312−1328.

(216) García de la Torre, J.; Huertas, M. L.; Carrasco, B. *J. Magn.Res.* **2000**, *147*, 138–146.

(217) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351–371.

(218) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714–2723.

(219) Ullmann, G. M.; Knapp, E.; Kostić, N. M. *J. Am. Chem. Soc.* **1997**, *119*, 42–52.

(220) Scanu, S.; Foerster, J. M.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2013**, *135*, 7681–7892.

(221) Hulsker, R. PhD thesis: Transient interactions between photosynthetic proteins, Leiden University, 2008, pp. 84–95.

(222) Worrall, J. A. R.; Liu, Y.; Crowley, P. B.; Nocek, J. M.; Hoffman, B. M.; Ubbink, M. *Biochemistry* **2002**, *41*, 11721–11730.

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

(224) Ubbink, M.; Bendall, D. S. *Biochemistry* **1997**, *36*, 6326–6335.

(225) Volkov, A. N.; Ferrari, D.; Worrall, J. A. R.; Bonvin, A. M. J. J.; Ubbink, M. *Protein Sci.* **2005**, *14*, 799–811.

(226) Volkov, A. N.; Bashir, Q.; Worrall, J. A. R.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010**, *132*, 1148–1195.

Appendix A: 1D-^{1}H NMR spectra of FKBP12 ligand titrations

General description: The proton assignments correspond to the numbers on the ligand structures shown in each figure. Signal intensities were scaled non-linearly for ease of viewing the shapes of the resonances. Asterisks indicate the protein/buffer/solvent resonances. Table A1 summarizes the details of the titration experiments shown in Figures A1-A8.

Table A1: Details of the titration experiments in Figures A1-A8.

Figure	ZB code	Initial [P] (μM)	Final [L] (mM)	Bound [L] at end point (%)	T_2 filter (ms)	K_D
A1	1	50	3.0	1.6	20	83 μM
A2	3	50	1.8	2.7	20	41 μM
A3	390	50	2.5	0.8	60	3.6 mM
A4	429	100	1.0	2.9	60	2.4 mM
A5	1051	100	1.0	n.d.[a]	0.5	n.d.
A6	1104	100	1.0	n.d.	60	n.d.
A7	1406	100	1.0	n.d.	0.5	n.d.
A8	1489	100	1.0	n.d.	20	n.d.

.[a] n.d.: not determined.
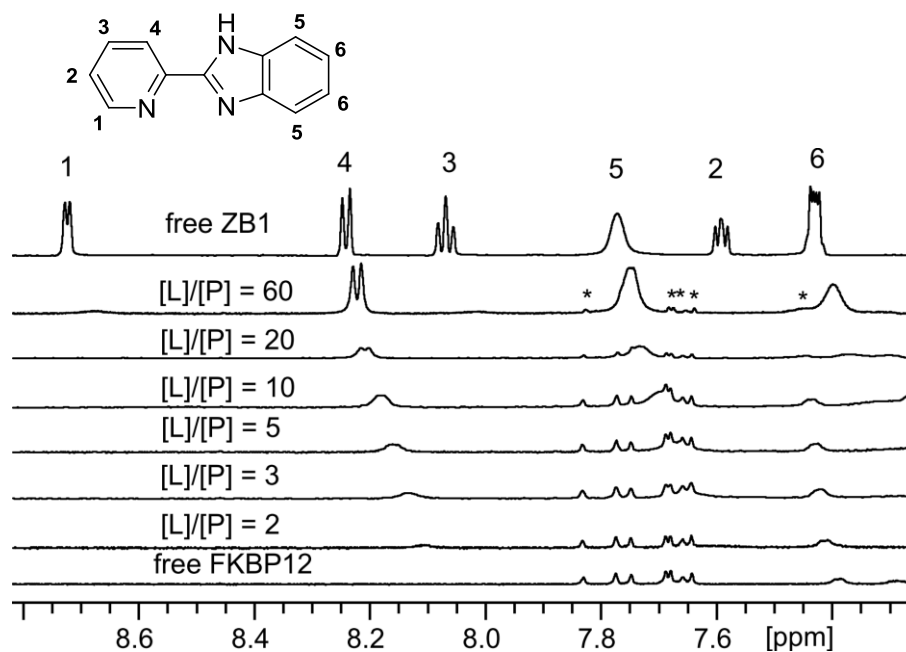
Figure A1: 1D-^{1}H NMR spectra of ZB1 titrated into wt FKBP12.
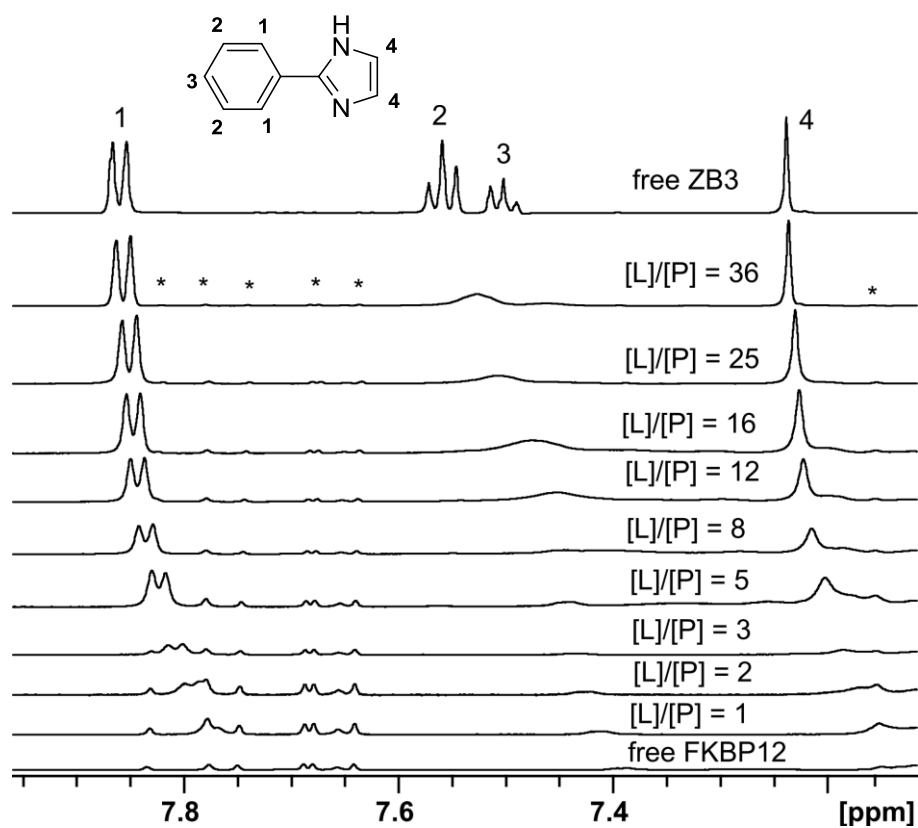
Figure A2: 1D-^{1}H NMR spectra of ZB3 titrated into wt FKBP12.
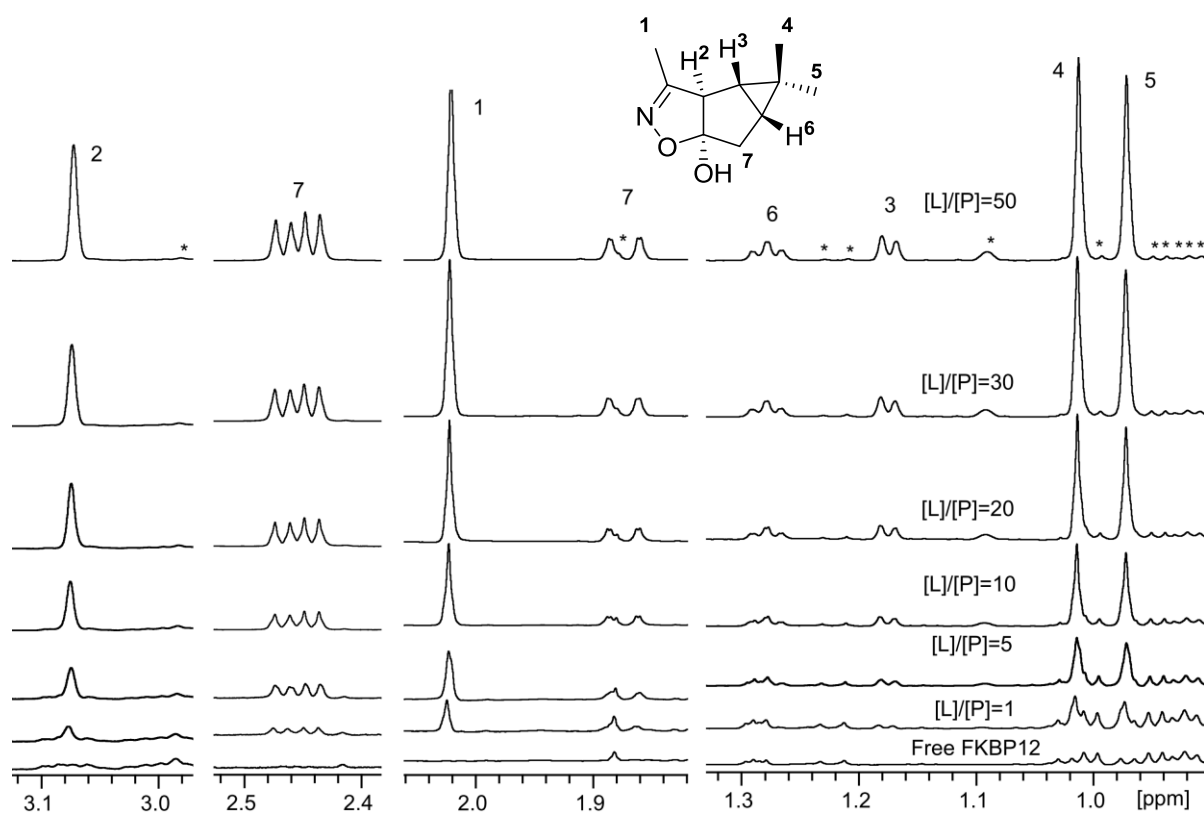
Figure A3: 1D-^{1}H NMR spectra of ZB390 titrated into wt FKBP12.
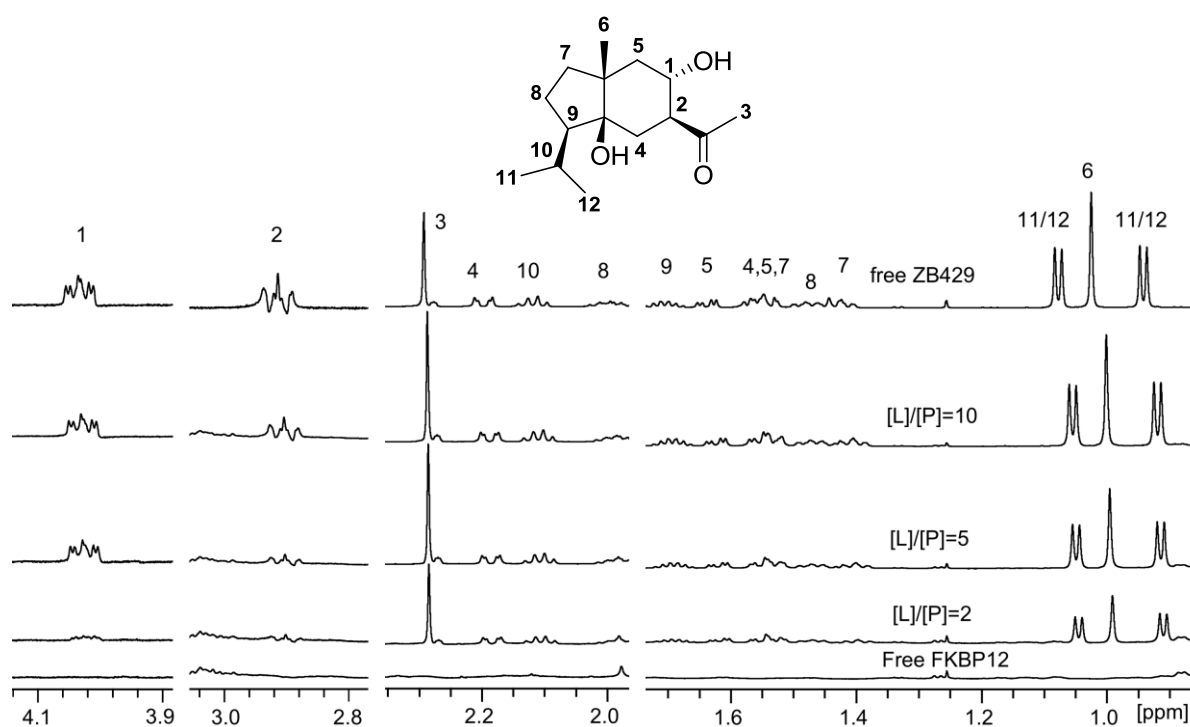
Figure A4: 1D-^{1}H NMR spectra of ZB429 titrated into wt FKBP12.
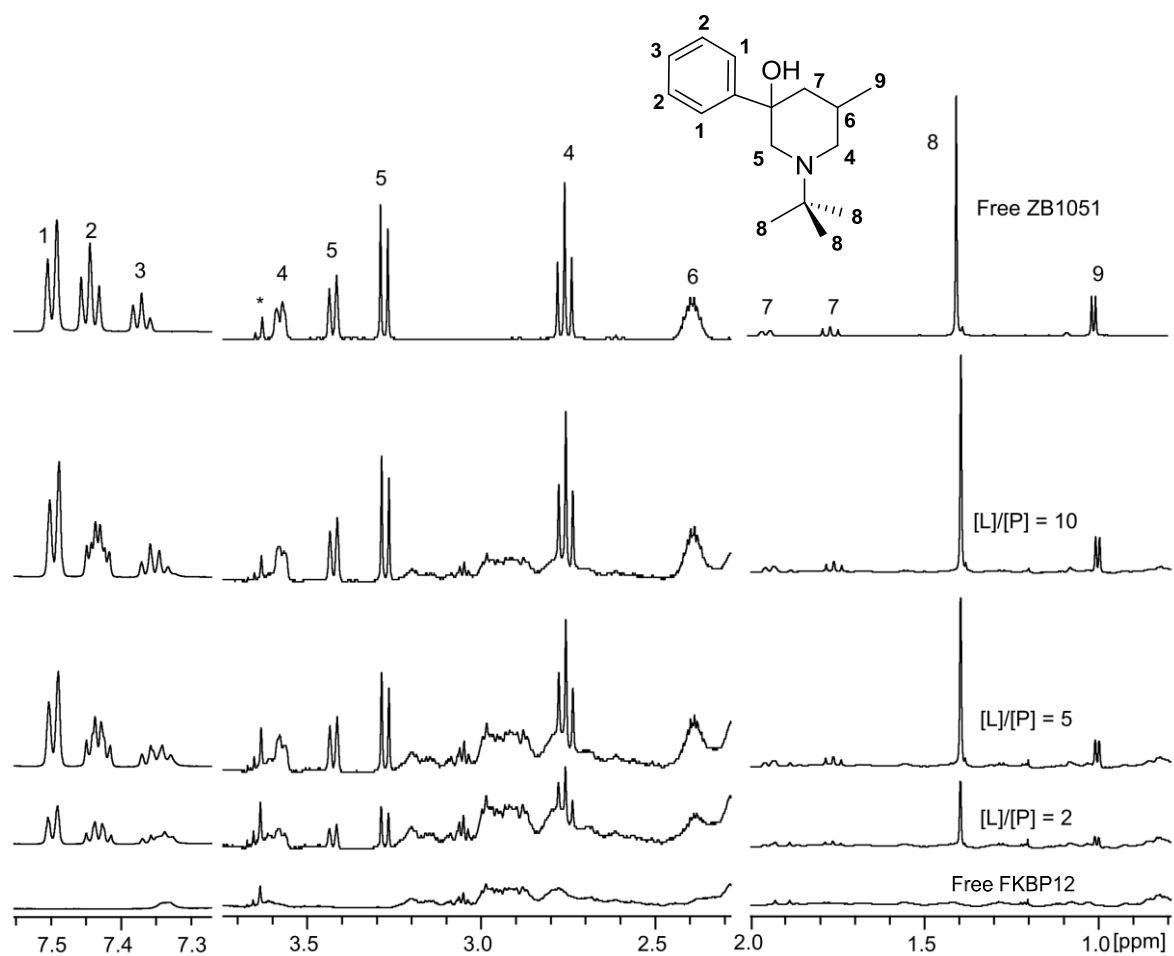
Figure A5: 1D-^{1}H NMR spectra of ZB1051 titrated into wt FKBP12.
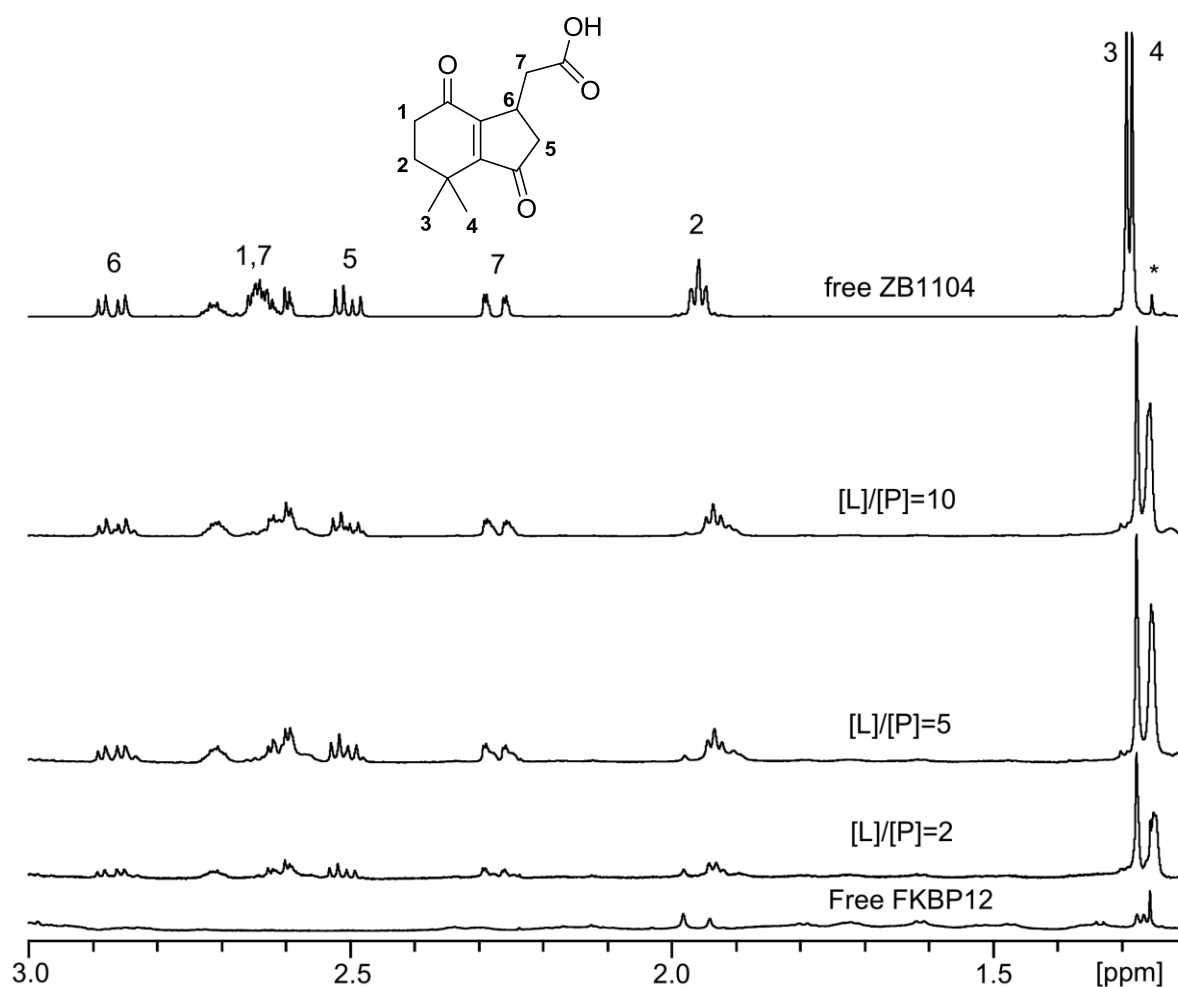
Figure A6: 1D-^{1}H NMR spectra of ZB1104 titrated into wt FKBP12. The proton assignments were based on prediction.
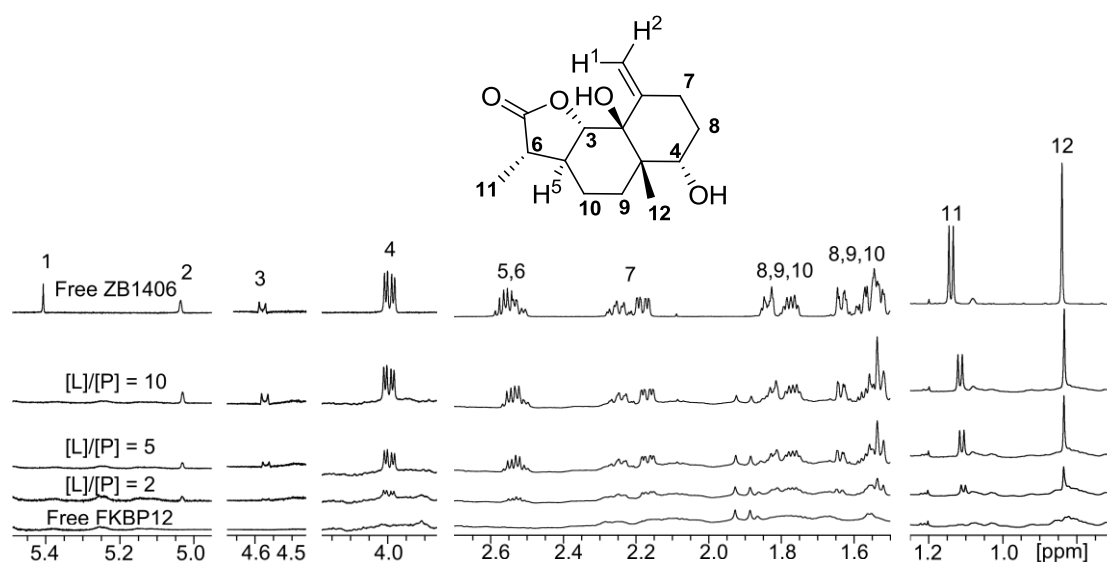
Figure A7: 1D-^{1}H NMR spectra of ZB1406 titrated into wt FKBP12. The proton assignments were based on prediction.
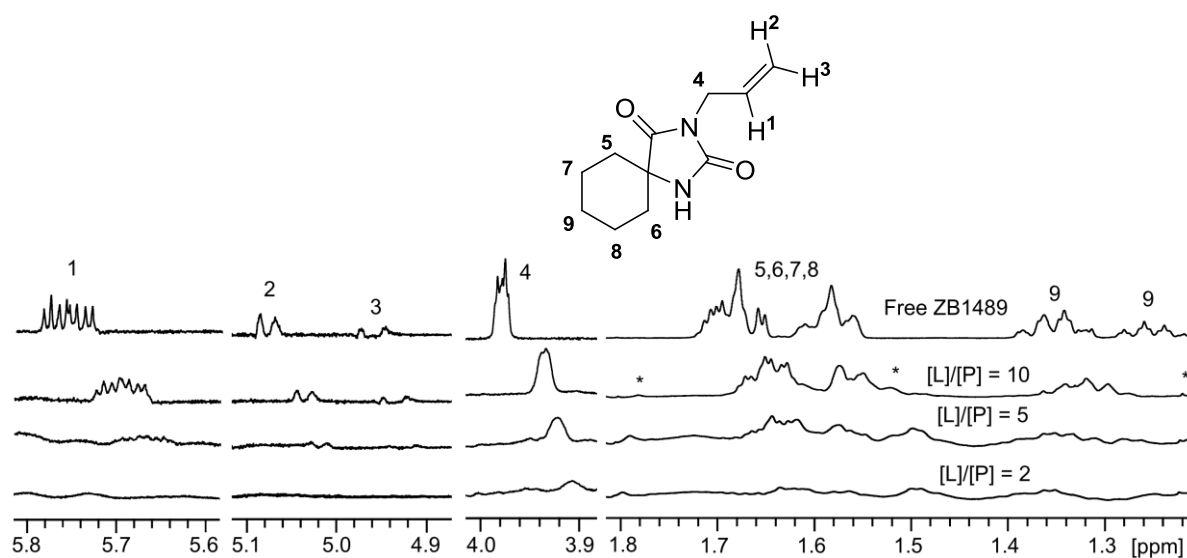
Figure A8: 1D-¹H NMR spectra of ZB1489 titrated into wt FKBP12. This compound showed complicated scalar coupling due to the aliphatic ring. The proton assignments were based on prediction.

Appendix B: XPLOR-NIH script for NOE- and PCS-based structure

calculations

```
display $DATE $TIME
! dock n ligand copies in FKBP12 using NOE restraints (pdb:1FKR14)
! metal positions and tensor directions placed on PDB of protein
! mutant 2 = 34-35 Cys
! mutant 3 = 44-47 Cys
! mutant 5 = 61-65 Cys, 22 Ala

!----------------DEFINE USER VARIABLES-----------------
 eval ($frbound2 = 0.015)               ! fraction ligand bound to mutant 2
 eval ($frbound3 = 0.009)               ! fraction ligand bound to mutant 3
 eval ($frbound5 = 0.019)               ! fraction ligand bound to mutant 5
 eval ($field = 600.13)                 ! magnetic field in MHz
 eval ($MUT2="Y")               ! data for mutant 2 Y/N
 eval ($MUT3="Y")               ! data for mutant 3 Y/N
 eval ($MUT5="Y")               ! data for mutant 5 Y/N
 eval ($metal2="Yb")            ! Ln in mutant 2: Yb or Tm
 eval ($metal3="Yb")            ! Ln in mutant 3: Yb or Tm
 eval ($metal5="Yb")            ! Ln in mutant 5: Yb or Tm
 eval ($chaxYbm2=8.7e-32)             ! chi axial in m^3 CLaNP-5 Yb mutant 2
 eval ($chrhoYbm2=3.4e-32)      ! chi rhombic in m^3 CLaNP-5 Yb mutant 2
 eval ($chaxYbm3=8.9e-32)             ! chi axial in m^3 CLaNP-5 Yb mutant 3
 eval ($chrhoYbm3=3.3e-32)      ! chi rhombic in m^3 CLaNP-5 Yb mutant 3
 eval ($chaxYbm5=7.8e-32)             ! chi axial in m^3 CLaNP-5 Yb mutant 5
 eval ($chrhoYbm5=2.7e-32)      ! chi rhombic in m^3 CLaNP-5 Yb mutant 5
 eval ($fres= 2)                ! first residue to be used
 eval ($lres= 107)             ! last residue to be used
 eval ($PRF= (1e30*1e6)/(12*3.14159)) ! factor in PCS equation
 eval ($f01=6.0)                ! force constant xpcs
 eval ($f02=6.0)                ! force constant xpcs
 eval ($f03=6.0)                ! force constant xpcs
 eval ($stdev1=0.04)            ! stdev for PCS
 eval ($stdev2=0.04)            ! stdev for PCS
 eval ($stdev3=0.04)            ! stdev for PCS
 eval ($a20=50)                 ! random distance max + 0 A
 eval ($a20a=360)               ! random rotation max +/-0.5*
 eval ($a05= 40000)             ! nr cycle
 eval ($a53= 300)               ! nr of steps
 eval ($a59= 10)                ! .pdb writing threshold
 eval ($a55= 300.0)             ! TBATH (temperatuur)
 eval ($rot_5ring="Y")          ! rorate ligand oxazole ring
 eval ($intac="N")              ! interactions to BB (Y) or BB + side chain (N)
 eval ($rigid="N")              ! dynamics mode = rigid (Y) or rigid+langevin" (N)
 eval ($pcs_r="Y")              ! docking with ligand PCS restraints
 eval ($noe_r="N")              ! docking with NOE restraints
 eval ($vdw_r="Y")              ! switch on/off vdw
 eval ($flex_noe="Y")           ! also docking with NOE from flexible loop
 eval ($lig_noe="Y")            ! include ligand intramolecular (transferred) NOE
 eval ($attraction_noe = "Y")   ! switch on/off attraction noe
 eval ($pcs_calc="Y")           ! back-calc PCS at end
 eval ($noe_calc="Y")           ! back-calc NOE distance at end
 eval ($opttens="N")            ! optimize tensor orienations iteratively
 eval ($kick=100)               ! # cycles per approach !100
 eval ($vdwmax=0.2)             ! starting rcon value, scaling of vdW energy
 eval ($lignr = 1)              ! number of ligand copies to be used
 eval ($genscale = 0.1)         ! general scale for noe
 eval ($noescale = 0.5)         ! scale real noes
 eval ($g01 = 20)               ! nr of cycles to get to max vdw in rigid docking
(should be 2<$g01< $kick) !20
!-----------------------------------------------------------------------
!------------OTHER VARIABLES; DO NOT CHANGE-----------------------------
!-----------------------------------------------------------------------
 eval ($vleckaxYbm2=$chaxYbm2/3.77E-35)    !chi axial in van Vleck units
 eval ($vleckrhYbm2=$chrhoYbm2/3.77E-35)   !chi rhombic in van Vleck units
 eval ($vleckaxYbm3=$chaxYbm3/3.77E-35)    !chi axial in van Vleck units
 eval ($vleckrhYbm3=$chrhoYbm3/3.77E-35)   !chi rhombic in van Vleck units
 eval ($vleckaxYbm5=$chaxYbm5/3.77E-35)    !chi axial in van Vleck units
 eval ($vleckrhYbm5=$chrhoYbm5/3.77E-35)   !chi rhombic in van Vleck units
 eval ($a01=1)               !cycle counter
 eval ($a02=$kick)           !counts cycles after last kick
 eval ($a02a=1) !counts approach step in which min ener was reached
 eval ($a02b=1)              !counts cycle number in current approach
 eval ($c01=0)               !counts number of saved structures
```

```
eval ($a58=30.0)          !fbeta, frictinal constant 30.0
eval ($a56=1.5)           !velocity factor! 1.5
eval ($a48=$cpu*1e4)      !($cpu*1e4)
set seed=$a48 end         !generate random number
eval ($a54 = 0.01)        !timestep in ps
eval ($f1 = 0)            ! van der waals scaling factor
eval ($kk2 = $kick/2)
!----------------------------------------------------------------------
!-------------------- READ STRUCTURE ----------------------------------
!----------------------------------------------------------------------
set echo=off end
set mess=off end
structure  @1FKR14.opt_lig.h.psf end
param  @1FKR14.opt_lig.h.par end
coor  @1FKR14.opt_lig.h.pdb
set echo=on end
set mess=on end

 parameter
        nbonds
                inhi=0.75
                ctofnb=5.0
                ctonnb=4.0
                repel=0.8
                rexp=2
                irex=2
                rcon=$f1
        end
 end

 parameters
    BOND   (segid "drg") (segid "drg") 400. TOKEN
    ANGLE  (segid "drg") (segid "drg") (segid "drg") 60. TOKEN
    IMPR   (segid "drg") (segid "drg") (segid "drg") (segid "drg") 50. TOKEN TOKEN
 end

 para
  nonb (name OO) 0.12  9.00      0.12  9.00
  nonb (name X)       0.0498   2.2272      0.0498   2.2272
  nonb (name Y)       0.0498   2.2272      0.0498   2.2272
  nonb (name Z)       0.0498   2.2272      0.0498   2.2272
  IMPR    OO   X   Z   Y 500.0 0 -54.7356
 end
!--------------------------rotate ligand oxazole ring------------------------
if ($rot_5ring="Y") then
 vector show (x) (resn DRG and name CAC)
 eval ($ix=$RESULT)
 vector show (y) (resn DRG and name CAC)
 eval ($iy=$RESULT)
 vector show (z) (resn DRG and name CAC)
 eval ($iz=$RESULT)
 vector show (x) (resn DRG and name CAD)
 eval ($jx=$RESULT)
 vector show (y) (resn DRG and name CAD)
 eval ($jy=$RESULT)
 vector show (z) (resn DRG and name CAD)
 eval ($jz=$RESULT)

 eval ($kx=$jx-$ix)
 eval ($ky=$jy-$iy)
 eval ($kz=$jz-$iz)

 coor rota SELE=(resn DRG and (name CAD or name OAE or name CAF or name HAG
   or name HAF or name CAK or name NAJ or name CAO or name CAL or name
   HAL or name HAM or name HAN or name HAR or name HAS or name HAQ))
      CENT=($jx $jy $jz) axis=($kx $ky $kz) -90.0 end
end if
!----------------------------------------------------------------------
!-------------------- DEFINE RESTRAINTS -------------------------------
!----------------------------------------------------------------------
!----------------DEFINE PCS----------------------
    ! ---- mutant 2, pcs in Hz, as observed ---
 vector do (store1=-2.5) (segid="drg" and name CAL)
 vector do (store1=-3.5) (segid="drg" and name CAO)
 vector do (store1=-2.5) (segid="drg" and name HAF)
 vector do (store1=-2.5) (segid="drg" and name HAG)
 vector do (store1=-3.1) (segid="drg" and name HAB)
```

```
 vector do (store1=-6.9) (segid="drg" and name HAA)
 vector do (store1=-9.4) (segid="drg" and name HAH)
 vector do (store1=-5.7) (segid="drg" and name HAI)
     ! ---- mutant 3, pcs in Hz, as observed ---
 vector do (store2=2.2) (segid="drg" and name CAL)
 vector do (store2=2.2) (segid="drg" and name CAO)
 vector do (store2=1.3) (segid="drg" and name HAF)
 vector do (store2=1.3) (segid="drg" and name HAG)
 vector do (store2=0.9) (segid="drg" and name HAB)
 vector do (store2=0.4) (segid="drg" and name HAA)
 vector do (store2=0.9) (segid="drg" and name HAH)
 vector do (store2=2.3) (segid="drg" and name HAI)
     ! ---- mutant 5, pcs in Hz, as observed ---
 vector do (store3=3.5) (segid="drg" and name CAL)
 vector do (store3=2.9) (segid="drg" and name CAO)
 vector do (store3=3.7) (segid="drg" and name HAF)
 vector do (store3=3.7) (segid="drg" and name HAG)
 vector do (store3=1.6) (segid="drg" and name HAB)
 vector do (store3=1.5) (segid="drg" and name HAA)
 vector do (store3=1.5) (segid="drg" and name HAH)
 vector do (store3=1.3) (segid="drg" and name HAI)

if ($pcs_r="Y") then
!------------------DEFINE CONDITIONS MUTANT 2-----------------------
 xpcs reset end
 xpcs nres 500 end

 if ($MUT2="Y") then
  xpcs
   class 2
   force $f01
   if ($metal2="Yb") then
     coeff $vleckaxYbm2 $vleckrhYbm2
   elseif ($metal2="Tm") then
     coeff $vleckaxTm $vleckrhTm
   end if
  end

  for $id in ID (store1)
  loop E
   vector show elem (store1) (ID $id)
   eval ($obs=$RESULT/($frbound2*$field))
   xpcs
     assign      (atom "tens" 1 OO)
           (atom "tens" 1 Z)
           (atom "tens" 1 X)
           (atom "tens" 1 Y)
           (ID $id) $obs $stdev1
   end
  end loop E
 end if
!-----------------DEFINE CONDITIONS MUTANT 3-----------------------

 if ($MUT3="Y") then
  xpcs
   class 3 !B
   force $f02
   if ($metal3="Yb") then
     coeff $vleckaxYbm3 $vleckrhYbm3
   elseif ($metal3="Tm") then
     coeff $vleckaxTm $vleckrhTm
   end if
  end

  for $id in ID (store2)
  loop A
   vector show elem (store2) (ID $id)
   eval ($obs=$RESULT/($frbound3*$field))
   xpcs
     assign      (atom "tens" 2 OO)
           (atom "tens" 2 Z)
           (atom "tens" 2 X)
           (atom "tens" 2 Y)
           (ID $id) $obs $stdev2
   end
  end loop A
 end if
```

```
!----------------DEFINE CONDITIONS MUTANT 5-----------------------

 if ($MUT5="Y") then
  xpcs
   class 5 !B
   force $f03
   if ($metal5="Yb") then
     coeff $vleckaxYbm5 $vleckrhYbm5
   elseif ($metal5="Tm") then
     coeff $vleckaxTm $vleckrhTm
   end if
  end

 for $id in ID (store3)
 loop G
   vector show elem (store3) (ID $id)
   eval ($obs=$RESULT/($frbound5*$field))
    xpcs
     assign      (atom "tens" 3 OO)
           (atom "tens" 3 Z)
           (atom "tens" 3 X)
           (atom "tens" 3 Y)
           (ID $id) $obs $stdev3
   end
  end loop G
 end if
end if
!----------------DEFINE REAL NOEs-----------------------
set echo=off end
set mess=off end
if ($noe_r="Y") then
@noe.xpl
end if
set echo=on end
set mess=on end
!----------------DEFINE ATTRACTION NOE-----------------------
if ($attraction_noe = "Y") then
noe
   class      attraction
   scale      attraction 0.5     !0.3 0.1
   sqconstant  attraction $genscale
   sqexponent  attraction 2
end

eval ($l1 = 1)
while ($l1 LE $lignr) loop G1
   noe
     assign (segid = "drg" and resi $l1 and name CAD) (name C2 and resn CM2) 13.0
13.0 5.0
     assign (segid = "drg" and resi $l1 and name HAT) (segid = "drg" and resi $l1
and name NAJ) 2.0  1.0  0.1  end
eval ($l1 = $l1 +1)
end loop G1

end if

!----------------------------------------------------------------------------
!--------------------DYNAMICS RUN & OUTPUT-------------------------------------
!----------------------------------------------------------------------------

 vector show (x) (name C2)
 eval ($xc2 = $RESULT)
 vector show (y) (name C2)
 eval ($yc2 = $RESULT)
 vector show (z) (name C2)
 eval ($zc2 = $RESULT)

 while ($a01 LE $a05) loop calc

  eval ($a02 = $a02 + 1)
  if ($a02 > $kick) then   ! ----------put ligand randomly away--------------
   coor init end           ! always use the same starting position of ligands
   coor @@1FKR14.opt_lig.h.pdb
   if ($rot_5ring="Y") then
    coor rota SELE=(resn DRG and (name CAD or name OAE or name CAF or name HAG
    or name HAF or name CAK or name NAJ or name CAO or name CAL or name
```

```
   HAL or name HAM or name HAN or name HAR or name HAS or name HAQ))
      CENT=($jx $jy $jz) axis=($kx $ky $kz) -90.0 end
  end if

   eval ($l1 = 1)
   while ($l1 LE $lignr) loop F2

     vector show ave (x)(segid="drg" and resi $l1)
     eval ($x1 = $RESULT)
     vector show ave (y)(segid="drg" and resi $l1)
     eval ($y1 = $RESULT)
     vector show ave (z)(segid="drg" and resi $l1)
     eval ($z1 = $RESULT)
     eval ($xr1 = $xc2 - $x1)
     eval ($yr1 = $yc2 - $y1)          ! move ligands to centre of protein
     eval ($zr1 = $zc2 - $z1)
     coor trans SELE=(segid="drg" and resi $l1) VECT=($xr1 $yr1 $zr1) end

     eval ($ang1=(0.5-RANDOM())*$a20a)
     eval ($vec1=(0.5-RANDOM())*3)
     eval ($vec2=(0.5-RANDOM())*3)
     eval ($vec3=(0.5-RANDOM())*3)  ! move and rotate ligands randomly away
     coor trans SELE=(segid="drg" and resi $l1) VECT=($vec1 $vec2 $vec3) DIST=50.0
end
     vector show ave (x)(segid="drg" and resi $l1)
     eval ($x1 = $RESULT)
     vector show ave (y)(segid="drg" and resi $l1)
     eval ($y1 = $RESULT)
     vector show ave (z)(segid="drg" and resi $l1)
     eval ($z1 = $RESULT)
     coor rota SELE=(segid="drg" and resi $l1)
            cent=($x1 $y1 $z1) axis ($vec3 $vec1 $vec2) $ang1 end
     eval ($l1 = $l1 + 1)
   end loop F2

  vector do (fbeta=$a58) (segid="drg")
  vector do (vx=$a56) (segid="drg")
  vector do (vy=$a56) (segid="drg")
  vector do (vz=$a56) (segid="drg")

  eval ($a02=1)
  eval ($miner=999999)
  !write coor FROM=MAIN SELE=(all) OUTPUT= temp2.pdb end
 end if

! --------------------dynamics run-----------------------------
  if ($intac="Y") then
   constraints interactions
    (segid="drg")
    (segid= "tens" or (segid "fkbp" and (name CA or name N or name C or name CB or
name HN or name O)))   end
  else
   constraints interactions (segid="drg")(segid "tens" or segid "fkbp") end
  end if

  constraints FIX=(not segid "drg") end

  if ($pcs_r="Y") then
   if ($vdw_r="Y") then
         if ($attraction_noe="Y") then
            flag exclude * include noe vdw xpcs end
         else
            flag exclude * include vdw xpcs end
         end if
   else
         flag exclude * include noe xpcs end
   end if
  else
   if ($vdw_r="Y") then
        flag exclude * include noe vdw end
   else
        flag exclude * include noe end
   end if
  end if

  para nbond rcon = $f1 end end
  if ($f1 < $vdwmax) then
```

```
     eval ($f1 = $f1 + $vdwmax/($g01-1))   !valid for next round
   end if
   display QQQQQ $f1

   display cycle $a01
   dynamics rigid
      dt=$a54
      group=(segid="drg" and resi 1)
      dynmode=TCOU !TCOU FREE LANG
      tbath=$a55
      nprint=50
      nstep=$a53
      NTRFRQ=0 !new for XPLOR vs 3.8
   end

if ($rigid="N") then
 if ($a02b > $g01) then
 !-----------local langevin dynamics--------------
 constraints
   interactions ((segid "fkbp" and not (name Ca or name C or name O or name N or
name
   HN or name C2)  and segid "drg" around 8.0) ) (segid "fkbp")
   interactions (segid "drg") (all)
 end

  constraints  !fix protein Bb + sc outside 8A from ligand + tens; ligand free
   FIX (not ((segid "fkbp" and not (name Ca or name C or name O or name N or name
   HN or name C2) and segid "drg" around 8.0) or (segid="drg") ))
   end

 display cycle $a01

 if ($pcs_r="Y") then
     if ($vdw_r="Y") then
          if ($attraction_noe="Y") then
           flag exclude * include noe vdw  bond angle dihedral impro xpcs end
          else
           flag exclude * include vdw  bond angle dihedral impro xpcs end

          end if
    else
          flag exclude * include noe bond angle dihedral impro xpcs end
     end if
 else
     if ($vdw_r="Y") then
          flag exclude * include noe vdw  bond angle dihedral impro end
    else
      flag exclude * include noe bond angle dihedral impro end
    end if
 end if
 energy end

 vector do ( fbeta = 6.657235 ) ((segid "fkbp" and not (name Ca or name C or name
O or name N or name HN or name C2)  and segid "drg" around 8.0) or segid "drg" )

 dynamics langevin
    timestep = 0.001
    nstep = 2000
    ilbfrq = 2000
    nprint = 200
    tbath = 300.
    iasvel = maxwell
    rbuf = 0.0
    origin = ( 0. 0. 0. )
 end

 minimize powell
  drop=1
  nprint=10
  nstep=50
 end

 end if
end if      !------------local langevin dynamics finished--------------

    !------------CALCULATE ENERGIES AFTER EACH CYCLE--------------
    constraints
```

```
  interactions (segid "drg") (segid "fkbp")
end
if ($pcs_r="Y") then!!MAKE SURE THIS MATCHES THE FLAG DURING THE RUN
    if ($vdw_r="Y") then
     if ($attraction_noe="Y") then
       flag exclude * include noe vdw xpcs end
     else
       flag exclude * include vdw xpcs end
     end if
    else
       flag exclude * include noe xpcs end
    end if
else
    if ($vdw_r="Y") then
       flag exclude * include noe vdw end  !xpcs
    else
       flag exclude * include noe end  !xpcs
    end if
end if
energy end

set disp=ener.dat end         !write energy values
if ($a01=1) then
    display file: ener.dat    $DATE $TIME
  if ($pcs_r="Y") then
        if ($vdw_r="Y") then
            if ($attraction_noe="Y") then
                display cycle   appr.step Etot Evdw Enoe Epcs
            else
                display cycle   appr.step Etot Evdw Epcs
            end if
        else
            display cycle   appr.step Etot Enoe Epcs
        end if
  else
        if ($vdw_r="Y") then
            display cycle   appr.step Etot Evdw Enoe
        else
            display cycle   appr.step Etot Enoe
        end if
  end if
end if

if ($pcs_r="Y") then
    if ($vdw_r="Y") then
        if($attraction_noe="Y") then
            display $a01    $a02a $a02 $ENER $VDW $NOE  $XPCS
        else
            display $a01    $a02a $a02 $ENER $VDW $XPCS
        end if
    else
        display $a01    $a02a $a02 $ENER $NOE  $XPCS
    end if
else
    if ($vdw_r="Y") then
      display $a01  $a02a $a02 $ENER $VDW $NOE
    else
      display $a01  $a02a $a02 $ENER $NOE
    end if
end if
set disp=OUTPUT end

if ($miner>$ENER) then
if ($a02b > $g01) then
coor copy end
 eval ($miner = $ENER)
 eval ($mincyc = $a01)
 eval ($minstep = $a02)
 if ($vdw_r="Y") then
        eval ($minvdw = $VDW)
 end if
 if ($noe_r="Y") then
     eval ($minnoe = $NOE)
 end if
 if ($pcs_r="Y") then
     eval ($minpcs = $XPCS)
     if ($attraction_noe="Y") then
```

```
                   eval ($minnoe = $NOE)
             end if
     end if
    end if
    end if
     !------------END OF APPROACH REACHED --------------
   if ($a02 = $kick) then
      if ($miner LE $a59) then
        !---write structure---
           eval ($c01 = $c01 + 1)
           eval ($a52="cycle_"+encode($c01)+".pdb")
       coor swap end
         write coord output=$a52 end
       !---write information about this cycle----
       set disp=coor.dat end !write details of step
       if ($c01=1) then
             display file: coor.dat   $DATE $TIME
          if ($pcs_r="Y") then
           if ($vdw_r="Y") then
            if($attraction_noe="Y") then
                 display file   cycle approach   app. step  Etot  Evwd  Enoe  Epcs
             else
                 display file    cycle approach   app. step  Etot  Evwd  Epcs
             end if
            else
                 display file    cycle approach   app. step  Etot  Enoe  Epcs
             end if
           else
             if ($vdw_r="Y") then
               display file cycle approach   app. step  Etot  Evwd  Enoe
             else
               display file cycle approach   app. step  Etot  Enoe
             end if
            end if
          end if
        if ($pcs_r="Y") then
          if ($vdw_r="Y") then
           if($attraction_noe="Y") then
             display $c01 $mincyc $a02a $minstep $miner $minvdw $minnoe $minpcs
           else
             display $c01 $mincyc  $a02a $minstep $miner $minvdw $minpcs
           end if
             else
               display $c01 $mincyc $a02a $minstep $miner $minnoe $minpcs
             end if
         else
             if ($vdw_r="Y") then
               display $c01  $mincyc    $a02a $minstep  $miner    $minvdw    $minnoe
                 else
               display $c01  $mincyc    $a02a $minstep  $miner    $minnoe
             end if
         end if
           set disp=OUTPUT end
       !---write file with PCS violations for this cycle---
          if ($pcs_calc="Y") then
           @pcs_calc.xpl
          end if
       !---write file with distances of atoms with NOEs---
           if ($noe_calc="Y") then
        @noe_calc.xpl
       end if
       set echo=on end
       set mess=on end
       end if
      eval ($a02a = $a02a + 1 )   !approach counter
      eval ($f1 = 0)   !set vdw back to zero
      eval ($a02b = 0) !cycle per approach
   end if

  eval ($a02b = $a02b + 1) !cycle per approach
  eval ($a01 = $a01 + 1)   !cycle counter
 end loop calc
display $DATE $TIME
!------------------------------------------------------------------
!---------------------FINISH----------------------------------------
!------------------------------------------------------------------
Stop
```

Appendix C: XPLOR-NIH script for PRE-based ensemble docking

```
! Energy minimization of TOAC pseudoatom position using PRE restraints
! Using Plastocyanin for docking TOAC on tetra-Lys peptide
! M. Ubbink 15.2.2013
! file: red_run.inp

 set mess=off end
 set echo=off end
!-------------------------------------------------------------------------
!---------------VARIABLES-------------------------------------------------
!-------------------------------------------------------------------------
 eval ($idi="plas")          ! SEGID of plastocyanin
 eval ($d1= 0.1)             ! lower bound CL3, real PRE
 eval ($d2= 0.1)             ! upper bound CL3, real PRE
! eval ($d3= 10.0)           ! target distance CL1
 eval ($d4= 7)               ! lower bound CL1, disappeared
 eval ($d5= 0)               ! upper bound CL1, disappeared
 eval ($d6= 15)          ! target distance CL2, unaffected
 eval ($d7= 0)               ! lower bound CL2, unaffected
 eval ($d8= 100.0)           ! upper bound CL2, unaffected
 eval ($d9= 0.5)             ! general scaling
 eval ($d10= 0.1)            ! scaling CL1 (disappeared) 1
 eval ($d11= 0.1)            ! scaling CL2 (not affected) 0.1
 eval ($d12= 0.1)            ! scaling CL3 (real PREs) 1
 eval ($d13= 100)            ! nr steps Powell energy minimization
 eval ($d15= 0.01)           ! scaling NOE to pull all the Toac to the protein
 eval ($tau_c= 5.54E-9)! tau-c in s
 eval ($fbeta=0.06)   !TC friction coefficient in psec-1
 eval ($run_nr= 1)    !cycle counter, $a01 in fkbp script
 eval ($kick=20)             ! # cycles per approach
 eval ($total_cycles= 11000)     ! nr runs EM (nr cycles),  $a05 in fkbp script
 eval ($a02=1)               !counts cycles after last kick
 eval ($min_app=1)    !counts approach step in which min ener
                               ! was reached, $a02a in fkbp script
 eval ($min_cycle=1) !counts cycle number in current approach, $a01 in fkbp script
 eval ($a20=40)              ! random distance max + 0 A
 eval ($a20a=30)             ! random rotation max +/-0.5*
 eval ($c01=0)               !counts number of saved structures, $c01 in fkbp script
 eval ($threshold=  4.00944*1.25)     ! .pdb writing threshold, $a59 in fkbp script

 eval ($frb=1)        ! fraction in which the peptide is bound to PoPc
 eval ($frl=1)               ! fraction of spin labelled TOAC peptide
 eval ($ens=6)               ! nr of TOAC molecules in ensemble
 eval ($t3=1)                ! first residue number of TOAC
 eval ($t4=$t3+$ens-1)
 eval ($a48=$cpu*1e4)   !($cpu*1e4)
 set seed=$a48 end          !generate random number
 eval ($Crd="Y")      ! Control loop Crd (Y/N)to kick TOAC after each approach
 eval ($ener_check="Y") ! switch on energy barrier (Y/N)
 eval ($checkener=2.1*$threshold)  ! energy barrier to prevent coordinate
explosion
 eval ($attraction_noe="Y") ! switch on attraction noe (Y/N)
 eval ($fix_CL2="Y")    ! fix experimental distance for CL2 (unaffected)
!-------------------------------------------------------------------------
!---------------PARAMETERS & STRUCTURE -----------------------------------
!-------------------------------------------------------------------------

 parameter
     @PoPc.cu.h.par
 end

 parameter
  NONBonded  TC      0.1200   3.3      0.1200   3.3
 end

 structure
     @PoPc.cu.h.psf
 end

 coordinates
     @PoPc.cu.h.pdb


!---------------define TOAC ------------------------------
```

```
topology
   residue TOAC
     atom TC type=C charge=0.000 mass 12.00 end
   end
end


segment
   molecule
     name=TOAC
     number=$ens
   end
end

para
 nbon
   repe = 0.65
   rcon = 60

 end
end

vector do (X=10.0) (resn TOAC and name TC)!
vector do (Y=10.0) (resn TOAC and name TC)!define initial TC coordinates
vector do (Z=10.0) (resn TOAC and name TC)!

!------------------------------------------------------------------------
!---------------RESTRAINTS ----------------------------------------------
!------------------------------------------------------------------------

 !DEFINE NMR DERIVED DISTANCES FOR TC-NH
 @restraints_PoPc_KKKKX_Kd93uM.xpl

 if ($attraction_noe="Y") then !
 NOE
   class CL4
   averaging CL4 cent
   potential CL4 square
   sqconstant CL4 0.01
   sqoffset CL4 0.0
   scale CL4 $d15
   sqexponent CL4 2
   ceil=10.0 !20
 end

eval ($t1 = $t3)
while ($t1 LE $t4) loop D ! set random coordinates for TOAC atoms
 noe
   assign (name TC and resi $t1) (name C2 and resn CM2) 20.0 10.0 10.0
 end
eval ($t1 = $t1 + 1)
end loop D
end if ! if ($attraction_noe="Y") then

!------------------------------------------------------------------------
!---------------RUN -----------------------------------------------------
!------------------------------------------------------------------------

 set echo=off end
 set mess=off end

 constraints fix (not name TC) end
 constraints interactions (name TC) (not name TC and not name C2) end
 flag exclude * include noe vdw end

 while ($run_nr LE $total_cycles) loop RUNS
   eval ($t1 = $t3)
   !coor orient sele=(segid "plas") end
   if ($Crd="Y") then
    if ($a02 = 1) then
     while ($t1 LE $t4) loop Crd      ! set random coordinates for TOAC atoms
      !-----------put TC at origin-----------------
      vector do (fbeta=0.025) (name TC and resi $t1)
      vector do (vx=0.1)(name TC and resi $t1)
      vector do (vy=0.1)(name TC and resi $t1)
      vector do (vz=0.1)(name TC and resi $t1)
      vector show ave (x) (name TC and resi $t1)
```

```
      eval ($x2= $RESULT)
      vector show ave (y) (name TC and resi $t1)
      eval ($y2= $RESULT)
      vector show ave (z) (name TC and resi $t1)
      eval ($z2= $RESULT)
      eval ($dist=sqrt($x2**2+$y2**2+$z2**2))
      coor trans SELE=(name TC and resi $t1) VECT=(-$x2 -$y2 -$z2) DIST=$dist end

      !------------ move TOAC randomly away---------------------

      eval ($vec1=0.5-RANDOM())
      eval ($vec2=0.5-RANDOM())
      eval ($vec3=0.5-RANDOM())
      coor trans SELE=(name TC and resi $t1) VECT=($vec1 $vec2 $vec3) DIST=$a20 end
      eval ($t1 = $t1 +1)
    end loop Crd

   minimize powell
     drop=10
     nprint=50
       nstep=$d13
       tolgradient=0.0001
   end

  end if ! if ($a02 = 1) then
  end if ! if($Crd="Y") then

  dynamics internal
    nstep = 500
    depred = 1
    etol = 1
    maxenergy = 100
    timestep = 0.002
    friction = $fbeta
    nsavc = 20
    fix = (segid $idi)
    group = (name TC and resi 1)
    group = (name TC and resi 2)
    group = (name TC and resi 3)
    group = (name TC and resi 4)
    group = (name TC and resi 5)
    group = (name TC and resi 6)
    tbath = 300
    nprint = 50
      NTRFRQ=0 !new for XPLOR vs 3.8
  end

  energy end

!-------------write all energy info------------------------

  set disp=ener.dat end          !write energy values
  if ($run_nr=1) then
   display ener.dat   $DATE $TIME
   display cycle      appr step Etot  Enoe Evdw
  end if
  display $run_nr     $min_app $a02    $ENER $NOE    $VDW
  set disp=OUTPUT end

!-------------check energy-----------------------------------
  if ($ener_check="Y") then
    if ( $a02 = 1 ) then
    if ($ENER>$checkener) then
      display ----------------------ENERGY TOO HIGH, GO TO NEXT CYCLE -----------
---------
      eval ($a02=0)
      eval ($min_app = $min_app + 1 ) !approach counter
     end if ! if ($ENER>$checkener) then
    end if !if ( $a02 = 1 ) then

    if ( $a02 = 2 ) then
     if ($ENER>$miner) then
      display ----------------------ENERGY TOO HIGH, GO TO NEXT CYCLE -----------
---------
      eval ($a02=0)
      eval ($min_app = $min_app + 1 ) !approach counter
     end if ! if ($ENER>$miner) then
```

```
      end if ! if ( $a02 = 2 ) then
   end if !if ($ener_check="Y") then
!-----------save energy info for currently lowest energy structure-------------

  if ($a02 = 1) then
    eval ($miner=999999)
  end if

  if ($miner>$ENER) then
    coor copy end
    eval ($miner = $ENER)
    eval ($mincyc = $run_nr)
    eval ($minstep = $a02)
    eval ($minvdw = $VDW)
    eval ($minnoe = $NOE)
  end if


!-----------END OF APPROACH REACHED --------------

!----------------------------------------------------------------------
!---------------ANALYSIS-----------------------------------------------
!----------------------------------------------------------------------

!-----------write the coordinates for lowest energy structure-------------

  if ($a02 = $kick) then

    if ($miner LE $threshold) then
     eval ($c01 = $c01 + 1)
     eval ($a52="cycle_"+encode($c01)+".pdb")
     coor swap end
     write coord sele=(name TC) OUTPUT=$a52 end

     !-----------save the PRE info for lowest energy structure-------------
     @pre_calc.xpl

     !-----------write the energy info for lowest energy structure-------------

     set disp=coor.dat end !write details of step
      if ($c01=1) then
            display file: coor.dat   $DATE $TIME
        display file      cycle approach  app.step   Etot Enoe Evwd violation
       end if
        display $c01 $mincyc   $min_app  $minstep  $miner    $minnoe    $minvdw
      $viol
      set disp=OUTPUT end


    end if ! if ($miner LE $threshold) then

    eval ($min_app = $min_app + 1 )   !approach counter
    eval ($min_cycle = 0) !cycle per approach
    eval ($a02 = 0)
  end if !  if ($a02 = $kick) then
  display ----------------------END OF RUN $run_nr----------------------

!----------------go to the next run----------------------------
  eval ($run_nr = $run_nr +1) ! run counter
  eval ($a02 = $a02 + 1)
  eval ($min_cycle = $min_cycle + 1) !cycle per approach


end loop RUNS
display $DATE $TIME
set echo=off end
set mess=on end
!-------------------------------------------------------------
!----------------------FINISH---------------------------------
!-------------------------------------------------------------
Stop
```

Appendix D: Backbone and side chain assignments of Zn-substituted PoPc

Residue	Cα	Cβ	H	N	Residue	Cα	Cβ	H	N
I1	59.73	40.89	-	-	H37	53.92	36.59	7.852	110.2
I1	59.76	38.60	-	-	N38	52.66	39.93	9.196	120.3
D2	52.86	42.05	7.793	126.2	I39	61.92	40.27	6.082	112.6
D2	53.10	41.58	8.310	111.2	V40	61.15	34.73	8.605	126.4
V3	60.62	34.86	8.810	124.1	F41	57.89	39.58	8.529	123.0
V3	60.64	35.08	8.773	124.0	D42	53.75	42.31	8.385	121.3
L4	54.68	42.58	8.963	126.0	E43	59.06	29.61	8.748	125.7
L5	53.18	41.42	8.626	119.4	D44	55.37	41.47	8.314	117.6
G6	44.37	0.00	7.816	116.2	S45	56.75	64.33	8.125	116.5
A7	50.76	20.87	7.995	125.1	I46	58.83	39.45	7.060	114.9
D8	57.51	39.98	8.935	119.9	P47	59.77	31.97	-	-
D9	53.42	39.59	7.627	115.3	S48	59.76	63.32	8.235	115.5
G10	44.71	0.00	8.189	108.7	G49	45.11	0.00	8.729	112.8
S11	60.35	63.09	8.046	115.8	V50	62.26	32.42	7.372	121.1
L12	53.16	39.11	8.682	125.9	D51	52.25	41.03	8.251	126.7
A13	50.06	22.39	7.351	124.6	A52	54.95	18.47	9.348	130.0
F14	58.35	40.45	9.077	121.2	S53	61.47	63.13	8.617	114.2
V15	59.31	34.20	8.823	121.5	K54	56.76	32.96	7.314	118.8
P16	63.63	35.03	-	-	I55	61.15	38.30	6.984	109.6
S17	59.21	64.62	8.160	107.7	S56	58.17	67.13	6.649	110.2
E18	55.26	32.27	7.500	119.9	M57	56.75	33.54	8.213	122.5
F19	56.00	39.07	7.632	119.0	M57	56.53	33.12	8.247	122.6
S20	56.74	66.30	8.592	115.3	S58	58.50	63.50	8.548	115.8
S20	56.67	66.27	8.626	115.3	E59	59.72	29.48	8.597	122.9
I21	59.62	42.71	8.716	116.5	E60	56.38	29.80	8.047	113.4
I21	59.70	43.03	8.716	116.5	D61	54.05	42.14	7.601	122.5
S22	56.78	63.29	8.228	116.8	L62	53.52	46.62	7.723	118.3
S22	56.99	63.17	8.406	117.4	L63	54.07	40.34	8.882	119.7
P23	64.51	31.46	-	-	N64	54.79	42.10	8.467	124.7
P23	64.39	31.48	-	-	A65	50.11	21.55	8.349	121.2
G24	44.97	0.00	9.216	113.3	K66	58.67	32.49	8.538	125.0
G24	44.95	0.00	9.095	113.9	G67	45.13	0.00	8.806	113.6
E25	56.99	31.06	7.655	121.7	E68	58.18	32.33	7.183	119.9
E25	56.60	30.65	7.801	123.1	T69	59.38	73.68	8.373	110.6
K26	56.16	33.67	8.040	122.3	T69	59.48	73.63	8.382	110.8
K26	56.19	33.98	7.981	120.4	F70	57.87	43.78	8.661	120.7
I27	60.34	39.64	8.867	125.6	F70	57.88	43.72	8.632	120.8
I27	60.65	39.11	8.858	125.0	E71	54.00	32.74	7.789	126.2
V28	61.64	31.39	8.468	126.3	E71	54.01	32.75	7.826	126.2
V28	61.63	31.14	8.437	126.8	V72	60.21	35.52	8.568	120.1
F29	57.18	39.29	9.042	126.5	V72	60.24	35.50	8.509	119.9
K30	54.83	35.19	9.173	122.8	A73	50.39	20.61	8.235	128.9
K30	54.88	35.11	9.138	122.6	L74	53.46	43.46	7.595	120.3
N31	54.46	38.45	9.229	126.8	L74	53.35	43.48	7.556	120.3
N32	56.50	42.50	9.016	129.2	S75	59.70	64.97	8.685	115.2
A33	52.28	22.05	8.647	119.8	N76	53.95	37.54	7.901	120.0
G34	47.51	0.00	7.750	116.5	K77	58.09	33.51	8.726	124.5
F35	52.80	36.98	5.720	114.5	G78	43.70	0.00	8.683	108.2
P36	62.47	36.82	-	-	G78	43.69	0.00	8.745	108.2

Residue	Cα	Cβ	H	N
E79	56.04	31.99	8.224	119.2
Y80	56.54	39.77	9.513	122.6
S81	58.36	64.76	8.268	118.7
F82	55.10	41.15	7.978	119.8
Y83	56.12	41.21	9.183	115.4
C84	56.22	29.46	7.804	121.2
S85	64.80	61.34	10.230	126.5
P86	65.14	30.36	-	-
H87	54.11	34.88	7.189	112.5
Q88	60.88	28.55	7.982	115.1
G89	46.30	0.00	8.964	108.5
A90	51.44	19.11	7.296	121.3
G91	45.35	0.00	7.829	110.4
M92	56.89	30.37	7.758	121.5
V93	58.74	35.98	8.016	121.5
G94	45.22	0.00	8.258	112.2
K95	55.78	36.67	8.394	119.7
V96	56.90	33.50	9.022	122.2
T97	61.77	70.28	8.244	124.4
V98	60.94	31.57	9.236	128.7
N99	55.42	42.04	8.648	132.1

Residue	H(sc1)	H(sc2)	N(sc)
N38	7.043	6.242	107.2
N64	7.631	6.865	113.5
Q88	6.731	6.189	110.8
N99	7.418	6.683	112.2

Appendix E: Backbone assignments of Zn-substituted DPc

| Residue	Cα	Cβ	H	N	Residue	Cα	Cβ	H	N
A1	51.87	20.70	-	-	A52	56.14	18.11	8.234	122.0
K2	55.70	35.28	8.579	123.2	S53	61.86	63.17	8.185	111.9
V3	61.23	35.98	8.568	125.4	E54	59.61	30.32	7.695	124.9
E4	55.79	30.83	9.265	126.7	L55	57.47	41.49	8.406	119.8
V5	61.75	31.83	9.060	125.4	K56	59.66	32.20	8.160	122.0
G6	43.85	-	8.015	112.6	A57	54.16	17.97	7.772	120.9
D7	52.58	43.17	6.938	109.9	A58	51.60	19.40	7.223	121.2
E8	59.14	29.44	8.915	116.9	S59	58.12	66.67	7.159	108.3
V9	62.03	31.73	7.415	113.5	M60	56.24	34.29	8.489	123.1
G10	45.51	-	7.841	108.9	D61	55.08	42.98	8.567	120.2
N11	54.45	39.01	9.108	119.1	E62	59.42	29.74	8.460	123.5
F12	51.90	34.75	8.662	125.6	N63	53.28	39.44	8.655	115.4
K13	54.33	38.05	6.459	117.4	D64	53.26	43.33	7.593	121.7
F14	57.37	41.32	8.332	119.6	L65	54.30	47.17	8.001	117.4
Y15	54.80	41.22	9.328	118.4	L66	55.37	44.15	8.308	121.3
P16	63.69	35.15	-	-	S67	56.81	67.73	8.003	115.1
D17	53.10	40.29	7.859	111.0	E68	59.52	29.62	8.688	118.2
S18	56.93	64.35	7.184	114.5	D69	55.86	41.60	7.757	115.1
I19	59.99	42.59	8.649	125.1	E70	53.79	30.18	7.251	120.2
T20	61.43	70.56	8.469	124.6	P71	65.22	33.45	-	-
V21	58.40	35.73	9.124	119.6	S72	56.90	66.60	7.904	114.3
S22	58.03	64.49	8.138	116.4	F73	57.84	44.44	9.029	125.5
A23	54.41	18.25	8.672	125.7	K74	54.81	34.76	7.569	127.5
G24	45.16	-	8.738	113.1	A75	50.40	23.60	8.753	125.6
E25	55.98	30.80	8.118	123.4	K76	55.85	34.50	8.355	122.0
A26	51.72	19.64	8.497	127.6	V77	62.23	33.58	8.862	128.0
V27	61.44	32.63	8.924	122.4	S78	59.58	64.54	9.329	123.0
E28	54.95	30.76	8.439	126.7	T79	61.79	68.79	8.145	123.5
F29	57.07	40.30	9.443	127.5	P80	63.84	32.46	-	-
T30	60.83	71.85	8.771	118.2	G81	44.78	-	8.537	110.4
L31	56.23	43.71	9.455	130.6	T82	62.25	71.00	7.628	115.2
V32	62.51	33.33	8.743	131.5	Y83	56.60	42.36	9.569	128.0
G33	43.85	-	8.279	112.5	T84	61.67	71.15	9.610	119.2
E34	57.38	31.12	8.006	114.3	F85	54.32	40.36	8.089	121.4
T35	63.05	69.04	7.865	117.7	Y86	56.42	41.47	8.616	115.1
G36	46.38	-	8.556	112.7	C87	57.18	29.85	7.105	120.1
H37	55.24	36.53	6.585	116.3	T88	69.99	67.19	9.995	126.6
N38	51.16	40.36	9.503	123.8	P89	65.03	31.09	-	-
I39	61.51	40.08	5.990	113.0	H90	57.96	32.67	7.459	114.7
V40	61.34	35.83	8.706	125.3	K91	61.58	29.21	7.903	128.0
F41	58.75	40.72	8.264	124.0	S92	60.28	62.36	-	-
D42	54.35	42.60	8.822	122.9	A93	51.33	18.54	7.202	123.7
I43	58.10	37.75	8.469	125.2	N94	54.62	36.96	8.195	114.0
P44	62.77	32.41	-	-	M95	56.66	31.08	7.492	119.7
A45	53.68	18.05	8.606	126.7	K96	54.22	37.51	7.369	125.5
G46	45.31	-	8.774	110.3	G97	44.21	-	8.212	112.7
A47	50.79	17.41	7.482	123.5	T98	61.09	71.56	8.074	112.9
P48	62.40	32.67	-	-	L99	53.01	46.81	9.640	128.8
G49	47.82	-	8.887	110.9	T100	62.59	69.82	9.384	126.3
T50	64.51	67.95	8.040	112.2	V101	60.99	33.30	9.301	128.9
V51	65.49	32.34	6.966	124.0	K102	57.42	34.75	8.780	133.5

Appendix F: Supplementary figures for Chapter 5
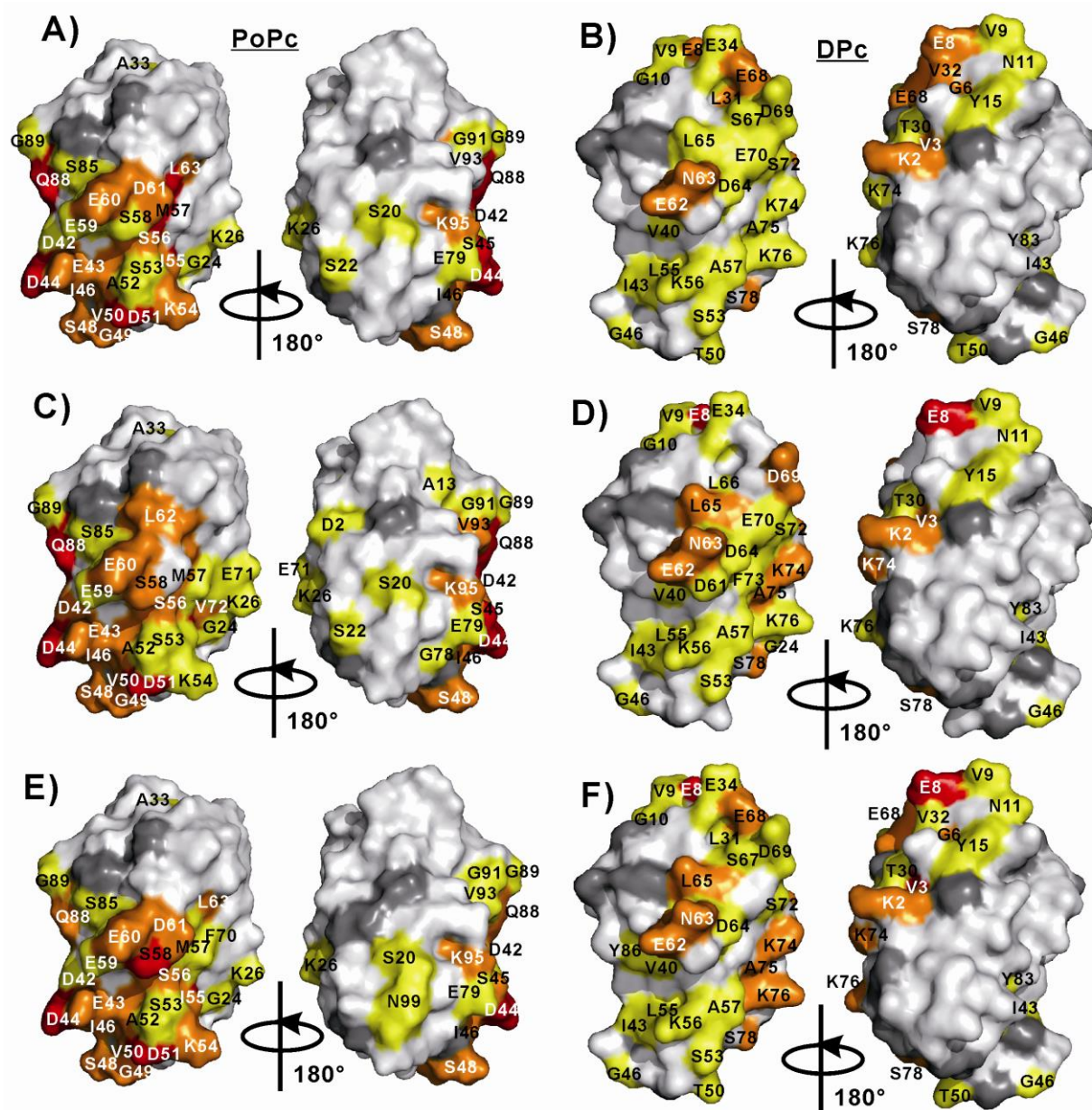
Figure F1: CSPs (extrapolated to 100% bound) mapped onto the protein surfaces from the binding of KKKKX (panels A and B), KKKKA (panels C and D) and AKKKK (panels E and F) to PoPc (left, PDB entry 1TKW[47]) and DPc (right, PDB entry 1KDI[43]). Color representations: red, $\Delta\delta_{ave} \geq 0.04$ ppm; orange, $0.04 > \Delta\delta_{ave} \geq 0.02$ ppm; yellow, $0.02 > \Delta\delta_{ave} \geq 0.01$ ppm; white, $\Delta\delta_{ave} < 0.01$ ppm; grey, no data or overlapping resonances.
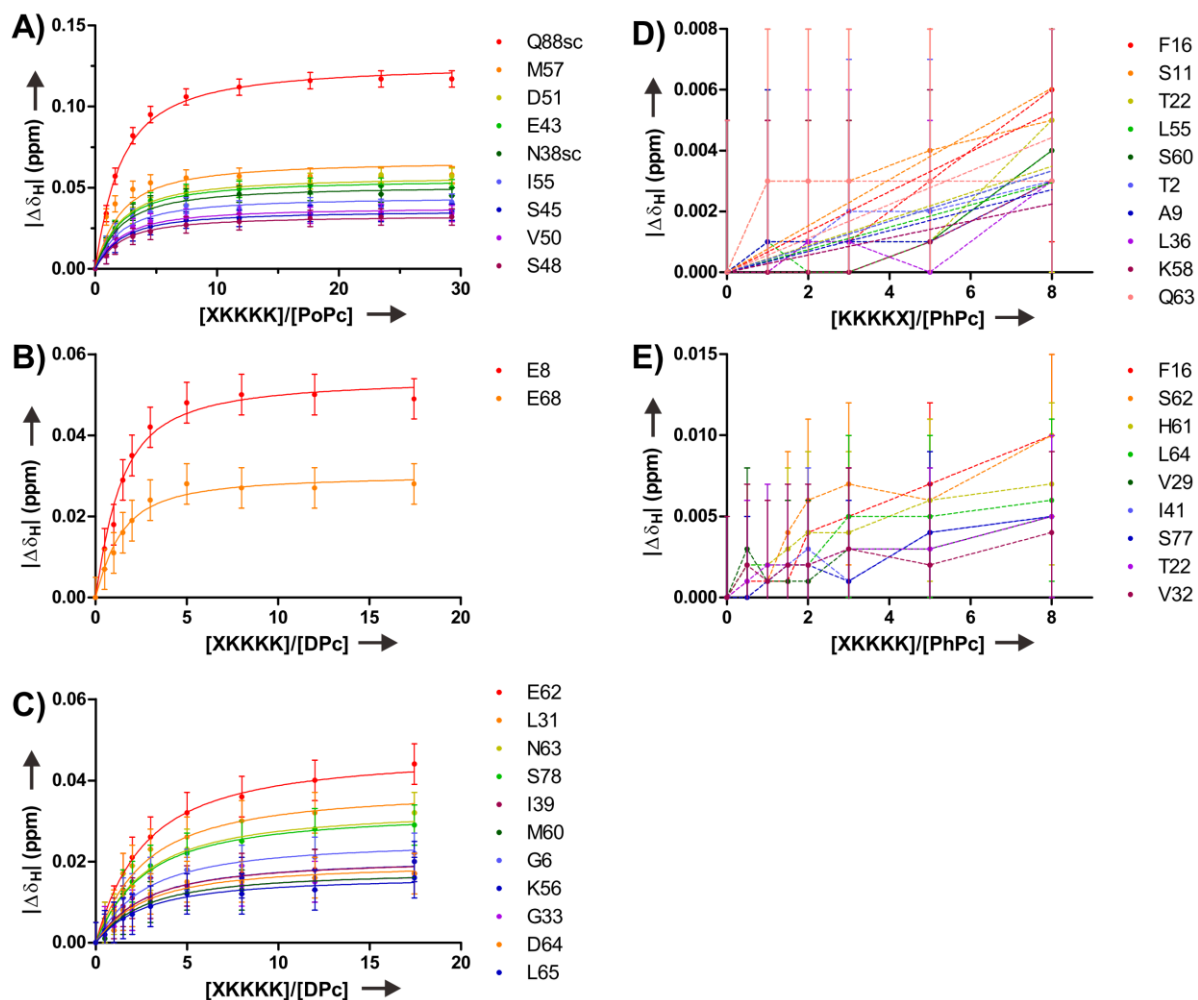
Figure F2: Chemical shift changes of Pcs resonances as a function of increasing [peptide]/[Pc] for peptides XKKKK and KKKKX. The dissociation constants of the corresponding peptides (Table 5.1) were obtained by simultaneous fitting to a 1:1 binding model for PoPc (solid lines) and by simulation for 2-site binding for DPc. (A) XKKKK with PoPc; (B) XKKKK with DPc, strong-binding residues; (C) XKKKK with DPc, weak-binding residues; (D) KKKKX with PhPc; (E) XKKKK with PhPc. The titration points for each residue in (D) and (E) are connected with dashed lines. Error bars represent ± 0.005 ppm.
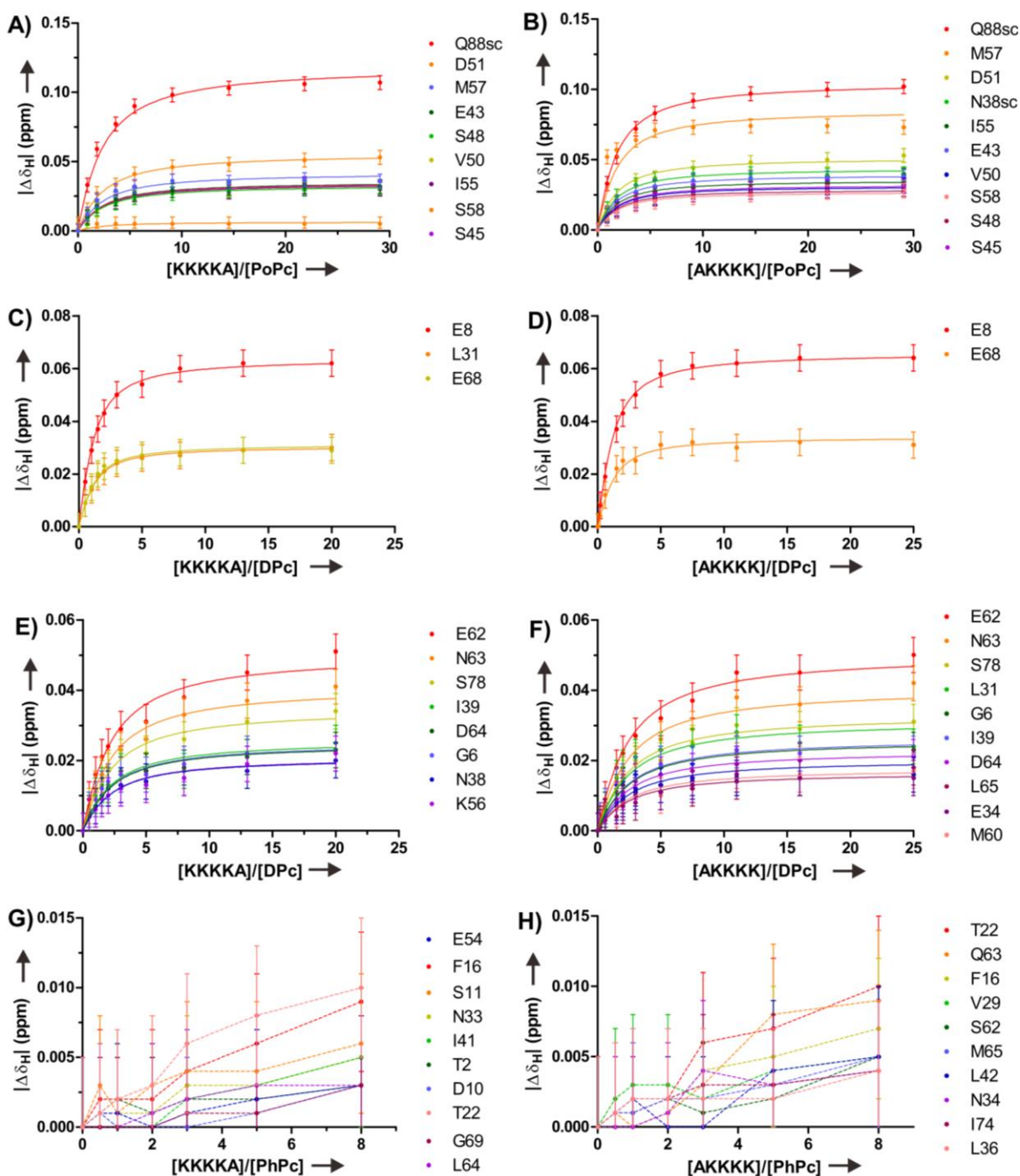
Figure F3: Chemical shift changes of Pcs resonances as a function of increasing [peptide]/[Pc] for peptides KKKKA and AKKKK. The residues which showed largest perturbations are shown. The dissociation constants of the corresponding peptides (Table 5.1) were obtained by simultaneous fitting to a 1:1 binding model for PoPc (solid lines) and by simulation for 2-site binding for DPc. (A) KKKKA with PoPc; (B) AKKKK with PoPc; (C) KKKKA with DPc, strong-binding residues; (D) AKKKK with DPc, strong-binding residues; (E) KKKKA with DPc, weak-binding residues; (F) AKKKK with DPc, weak-binding residues; (G) KKKKA with PhPc; (B) AKKKK with PhPc. The titration points for each residue in (G) and (H) are connected with dashed lines. Error bars represent ± 0.005 ppm.
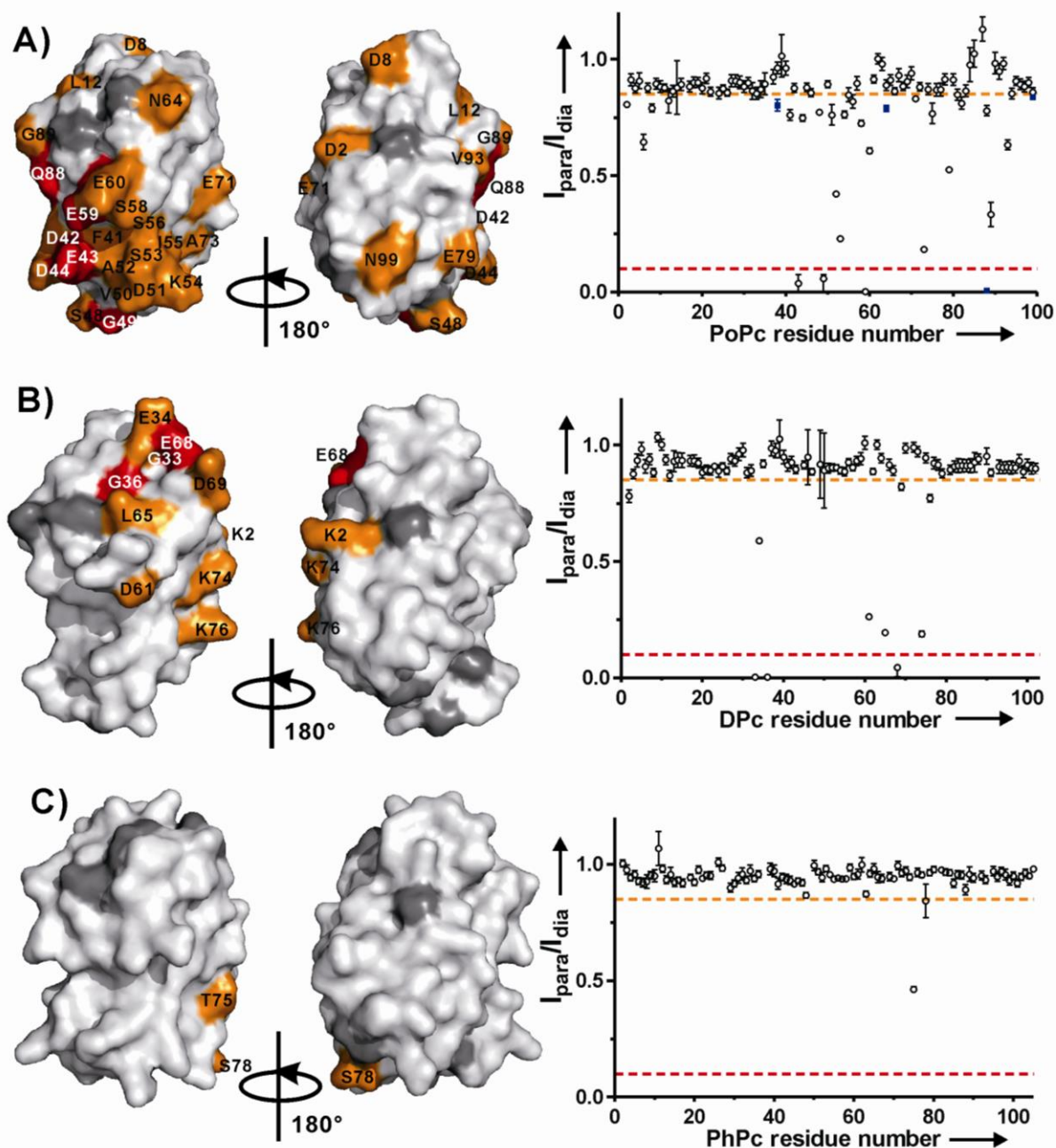
Figure F4: PRE effects in Pc-XKKKK complexes. Left: PRE maps of PoPc (A, PDB entry 1TKW[47]), DPc (B, PDB entry 1KDI[43]) and PhPc (C, PDB entry 2Q5B) bound to XKKKK peptide, color-coded on surface models of Pc: red, $I_{para}/I_{dia} < 0.1$; orange, $0.1 \leq I_{para}/I_{dia} < 0.85$; white, $I_{para}/I_{dia} \geq 0.85$; grey, prolines, unassigned, and overlapping resonances. Right: relative [^{1}H,^{15}N]-HSQC intensities of amides PoPc (A), DPc(B) and PhPc(C) in the complex with TOAC-containing peptides. For PoPc, the side chains are also included (blue squares). The dashed horizontal lines indicate $I_{para}/I_{dia} = 0.85$ (orange lines) and 0.1 (red lines). The error bars denote 2× standard deviations, derived from spectral noise levels using standard error propagation procedures.
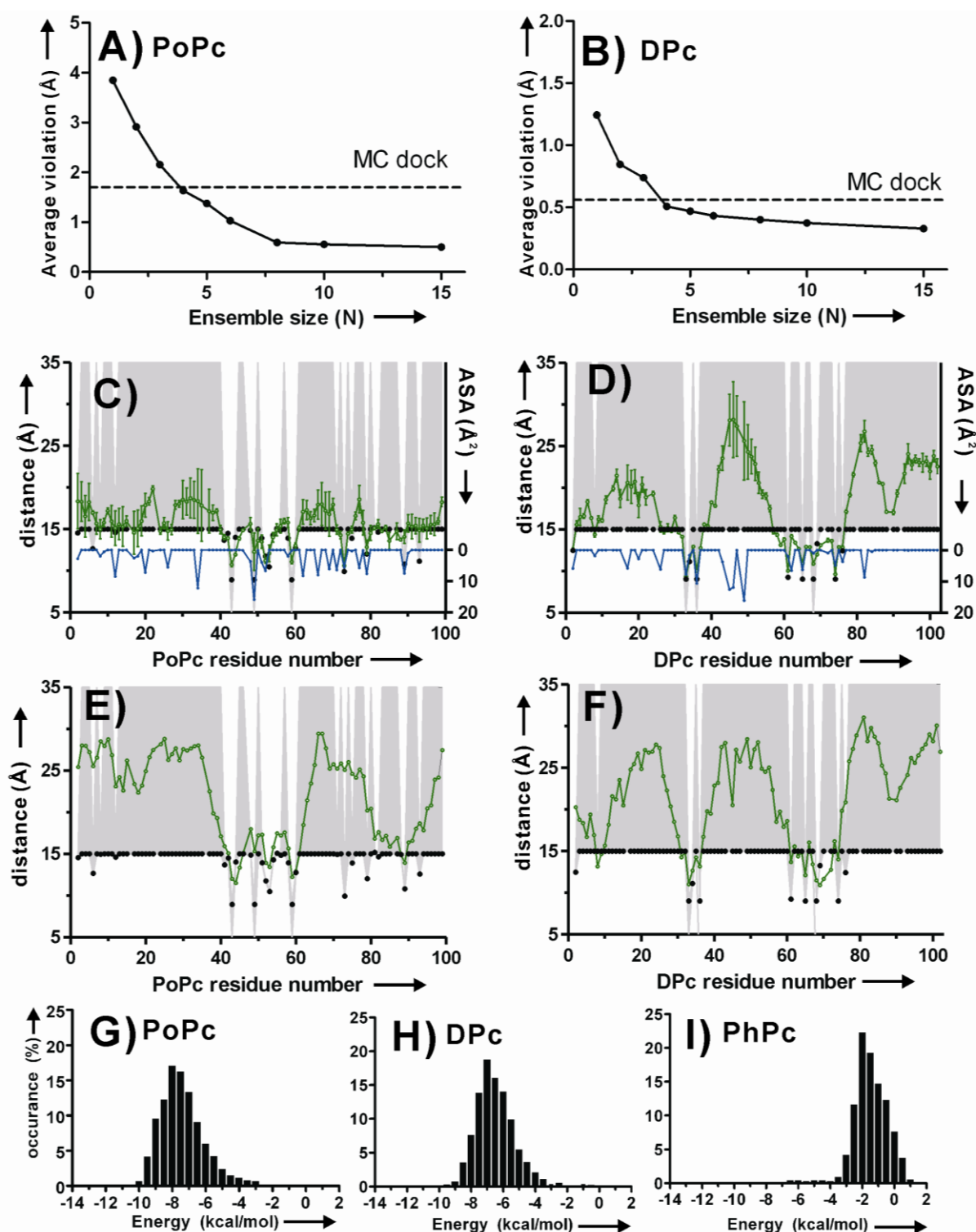
Figure F5: (A-B) Averaged distance violations against number of XKKKK peptides (N=1-6,8,10,15) in the ensemble docking for PoPc (A) and DPc (B). (C-D) Correlation of experimental distances (black dots) and back-calculated average distances (green circles with connecting lines) from the ensemble docking (N=6) of XKKKK bound to PoPc (C) and DPc (D). The average distances from the 20 lowest-energy solutions of the PRE driven ensemble docking are shown as black circles connected by black lines with error bars representing the standard deviation. Right y axes show the accessible surface area (ASA) of each amide. Grey areas indicate the error margins of the

experimental distances. (E-F) Comparison of experimental distances (black dots) and back-calculated average distances (green dots with connecting lines) between Pc amides and the 2000 ensembles of peptide paramagnetic oxygen atoms from MC simulations for PoPc (E) and DPc (F). Grey areas indicate the error margins of the experimental distances. (G-I) Histograms showing the energy distribution of 2000 ensembles from MC simulations: (G) PoPc-XKKKK, (H) DPc-XKKKK, (I) PhPc-XKKKK.

Summary

The work described in this thesis focuses on the application of various NMR techniques to the study of interactions between proteins and small molecules and proteins and peptides, including the well-established classical NMR approaches, and the recently developed paramagnetic NMR methods. Computational tools have been used to complement and visualize the experimental data.

Chapter 1 presents an overview of NMR techniques for studying protein-ligand interactions. Due to its superior sensitivity in detecting weak interactions, NMR is the most used technique for fragment-based drug screening and the power of NMR is not restricted to primary screening only.

The work presented in Chapter 2 gives detailed information on the characterization of fragment hits binding to FKBP12 using 1D-^{1}H NMR. Recent advances in paramagnetic NMR spectroscopy have evolved a wide range of applications in studying protein-protein complexes but few cases of protein-ligand complex studies have been reported. To look for suitable ligands used in a paramagnetic NMR study (Chapter 4), fifteen FKBP12 ligands with diverse binding affinities and chemical structures were selected for the pilot study. The suitabilities of the ligands were assessed based on their binding affinity, chemical structure, isomeric purity, and changes of chemical shifts upon binding to the target protein. Based on the characterization of the bound ligands, the general rules of an ideal ligand were established as: (1) the ligand must be in fast exchange on the NMR time scale; (2) must contain sufficient number of proton resonances that are well-separated in the 1D-^{1}H NMR spectrum; (3) must be isomerically pure; (4) must show clear and measurable chemical shift changes upon binding to the protein. The binding pose of the most suitable ligand was subsequently characterized in detail in Chapters 3 and 4.

Chapter 3 describes the determination of the binding pose of a small molecule bound to FKBP12 based on NOE restraints. Chemical shift perturbation (CSP) analysis indicated that a large region of the protein, spanning both binding sites 1 and 2 on FKBP12, was affected either directly or indirectly by ligand binding. The CSP map therefore presents a challenge to identify the actual binding site. Thus, intermolecular NOE restraints acquired from standard isotope filtered/edited NOESY experiments were used to determine the binding site and ligand binding mode. Although not all NOE restraints were satisfied in the final structure, due to dynamics in the complex, the results show a promising hydrogen bond network that has also been found in other FKBP12 ligands.

Chapter 4 presents an orthogonal NMR methodology to determine the structure of the above-mentioned complex based on paramagnetic NMR pseudocontact shifts (PCSs). The two-armed lanthanide binding tag, CLaNP5, was attached on three different double cysteine variants of FKBP12. The results indicate that it is possible to identify the ligand binding site and obtain a low resolution structure using bound ligand PCSs from simple 1D-^{1}H NMR spectra. Due to the rigid structure of CLaNP5, optimization of the tensor frame and the lanthanide position was not necessary but could improve the quality of the structure. Therefore, the methodology can be particularly valuable for studying proteins that are very large or difficult to be labeled isotopically.

The weak interaction and dynamic behavior of transient encounter complexes is challenging to study using existing experimental tools. The binding partners in such complexes present in nature, mostly being proteins, have low specificity and can interact with multiple partners in cascades. Chapter 5 describes the characterization of dynamics in artificial encounter complexes formed by plastocyanins and short charged peptides. Using CSP and PRE NMR spectroscopy, the dynamics in the encounter complexes were visualized and compared with the results from electrostatic Monte Carlo (MC) simulations. The highly similar CSP maps and the small shifts among different complexes strongly suggest a high degree of dynamics. In addition, the scattered PREs indicate the presence of multiple orientations. The overall results suggest that the complexes have multiple orientations and are dominated by electrostatics, and evidence for weak short-range interactions is provided by the PRE data.

Samenvatting

Het werk beschreven in dit proefschrift richt zich op de toepassing van diverse kernspinresonantie (*Nuclear Magnetic Resonance, NMR*) technieken voor het bestuderen van de interactie tussen eiwitten en kleine moleculen en tussen eiwitten en peptiden. Er is gebruik gemaakt van zowel klassieke NMR-technieken als recent ontwikkelde paramagnetische NMR methoden. Computeranalyses zijn gebruikt om de experimentele data aan te vullen en te visualiseren.

In Hoofdstuk 1 wordt een overzicht van NMR-technieken gegeven die gebruikt worden voor het bestuderen van eiwit-ligand interacties. Vanwege zijn superieure gevoeligheid voor de detectie van zwakke bindingen, is NMR de meest gebruikte techniek voor fragmentgebaseerde geneesmiddelen ontwikkeling, maar de kracht van NMR is niet beperkt tot alleen dit soort onderzoek.

Het werk gepresenteerd in Hoofdstuk 2 geeft gedetailleerde informatie over de karakterisatie van fragmenten die binden aan FKBP12, door gebruik te maken van 1D-^{1}H NMR. Recente ontwikkelingen in paramagnetische NMR-spectroscopie hebben geleid tot een breed spectrum van toepassingen voor het bestuderen van eiwit-eiwit complexen, maar er zijn weinig studies van eiwit-ligand complexen gerapporteerd. Om geschikte liganden te vinden voor toepassing van paramagnetische NMR-methoden (Hoofdstuk 4), zijn 15 FKBP12 liganden met verschillende bindingsaffiniteiten geselecteerd voor een eerste onderzoek. De geschiktheid van de liganden is bepaald op basis van bindingsaffiniteit, chemische structuur, isomere zuiverheid en veranderingen van chemische verschuivingen (*chemical shifts*) bij binding aan het eiwit. Op basis van de gevonden resultaten zijn de algemene regels voor een ideale ligand als volgt: (1) de ligand moet zich in snelle uitwisseling (*fast exchange*) bevinden op de NMR-tijdschaal; (2) de ligand moet een voldoende aantal goed gescheiden proton resonanties vertonen in het 1D-^{1}H NMR spectrum; (3) de ligand moet isomeer-zuiver zijn; (4) de ligand moet een duidelijke en meetbare chemische verschuiving vertonen bij binding aan het eiwit. De bindingspositie van het meest geschikte ligand is vervolgens gedetailleerd gekarakteriseerd, zoals beschreven in Hoofdstukken 3 en 4.

Hoofdstuk 3 beschrijft de bepaling van de bindingspositie van een klein molecuul gebonden aan FKBP12, gebaseerd op NOE (*Nucleaire Overhauser Effect*) gegevens. Analyse van de verstoringen in chemische verschuiving (*Chemical shift perturbation, CSP*) toonden aan dat een groot gebied van FKBP12, waaronder de bekende bindingsplaatsen 1 en 2, direct of

indirect wordt beïnvloed door binding van ligand. De CSP 'kaart' geeft daarom geen inzicht in de daadwerkelijke bindingsplek. Derhalve zijn NOEs, verkregen uit standaard isotoop gefilterde/gekoppelde NOESY-experimenten, gebruikt om de bindingsplaats en bindingsoriëntatie van het ligand te bepalen. De uiteindelijke structuur is niet in overeenstemming met alle NOEs, wat vermoedelijk wordt veroorzaakt door dynamiek van het ligand in de bindingsplaats. Toch is het resultaat waarschijnlijk betrouwbaar, mede omdat een waterstofbrugnetwerk aanwezig is, welk ook is gevonden bij andere FKBP12 liganden.

Hoofdstuk 4 presenteert een orthogonale NMR methode om de structuur van het hierboven genoemde complex te bepalen, gebaseerd op paramagnetische NMR *pseudocontact shifts* (PCSs). CLaNP-5, een eiwitprobe met twee armen welke een lanthanide-ion kan binden, is bevestigd aan drie verschillende dubbel-cysteïne mutanten van FKBP12. De resultaten tonen aan dat het mogelijk is de bindingsplaats van een ligand te identificeren en een structuur met lage resolutie te verkrijgen door gebruik te maken van ligand-PCSs verkregen uit simpele 1D-^{1}H NMR spectra. Door de rigide structuur van CLaNP5 was optimalisatie van het *tensor frame* en de positie van de lanthanide niet noodzakelijk, al leidde dit wel tot een hogere kwaliteit van de structuur. Om deze reden kan deze methode in het bijzonder waardevol zijn voor het bestuderen van grotere eiwitten of eiwitten die moeilijk te verrijken zijn met isotopen.

De zwakke interactie en het dynamisch gedrag van kortstondige complexen zijn een uitdaging om te bestuderen met de huidige experimentele hulpmiddelen. De bindingspartners van deze in de natuur voorkomende complexen, meestal eiwitten, hebben een lage specificiteit en kunnen opeenvolgende interacties vertonen met meerdere partners. Hoofdstuk 5 beschrijft de karakterisatie van de dynamiek in een kunstmatig kortlevend complex gevormd door plastocyanine en korte, geladen peptides. Door gebruik te maken van *CSP* en *PRE* NMR spectroscopie is de dynamiek van de kortstondige complexen gevisualiseerd en vergeleken met de resultaten van elektrostatische *Monte Carlo* (MC) simulaties. De zeer vergelijkbare CSP 'kaarten' tussen verschillende complexen evenals de kleine verschuivingen, suggereren een hoge mate van dynamiek. Bovendien tonen de verspreidde PRE's de aanwezigheid aan van meerdere oriëntaties. De algehele resultaten suggereren dat de complexen uit meerdere oriëntaties bestaan en door elektrostatische interacties worden gedomineerd. Tevens zijn er aanwijzingen voor zwakke korte-afstandinteracties op grond van de PRE data.

List of Publications

J.-Y. Guan, P. H. J. Keizers, W.-M. Liu, F. Löhr, E. Heeneman, S. P. Skinner, H. Schwalbe, M. Ubbink and G. Siegal. **Small molecule binding sites on proteins established by paramagnetic NMR spectroscopy.** *J. Am. Chem. Soc.*, **2013**, *135* (15), 5859–5868

J.-Y. Guan, J. M. Foerster, J. W. Drijfhout, M. Timmer, A. Blok, G. M. Ullmann and M. Ubbink. **An ensemble of rapidly interconverting orientations in electrostatic protein-peptide complexes characterized by NMR spectroscopy**. *Manuscript submitted to ChemBioChem.*

C. H. Röhrig, C. Loch, J.-Y. Guan, G. Siegal and M. Overhand. **Fragment-based synthesis and SAR of modified FKBP ligands: influence of different linking on binding affinity.** *ChemMedChem* **2007,** 2, 1054–1070.

Curriculum vitae

Jia-Ying Guan (官佳穎) was born on 1 December 1980 in Yi-Lan, Taiwan. She obtained her early education in the beautiful countryside of Yi-Lan and the technology town of Hsin-Chu, and then completed her bachelor in chemistry from National Taiwan University in Taipei.

After spending a few years in Taipei, she decided to take an adventure in the other side of the world, a country that has more bikes than people– the Netherlands. She did internships in the bioorganic synthesis group and the protein chemistry group in the Leiden Institute of Chemistry. The topics of her internships include fragment-based synthesis of FKBP12 ligands, synthesis of bacterial teichoic acid subunits, and comparison of various NMR screening techniques. In 2008 she completed her master degrees in chemistry from Leiden University, with specialization in Biological Chemistry and Design and Synthesis.

In 2009 she started her PhD research in the Protein Chemistry group under the supervision of Prof. Dr. Marcellus Ubbink and Dr. Gregg Siegal. The research focused on the applications of paramagnetic NMR in protein-ligand studies. Part of the work described in this thesis was presented at the International Conference on Magnetic Resonance in Biological Systems (ICMRBS) 2012 in Lyon, France and Annual Dutch Meeting on Chemical Sciences in Veldhoven in 2012.

Acknowledgements