universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Paramagnetic NMR studies of IDPs"

verfasst von / submitted by

Agathe Vanas, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Univ.-Prof. Dr. Robert Konrat

Acknowledgements

I would like to thank all the people who made it possible for me to pursue my studies and complete this work. First and foremost I would like to express my gratitude to my family and friends, particularly my mother and my sister who have always stood behind me and supported me in my endeavors.

Special thanks go to Robert Konrat who first inspired me to dive into the world of structural biology and then nurtured the passion I had found by giving me the opportunity to work in his laboratory and ultimately enabling this work. Another person who has earned my personal gratitude is Dennis Kurzbach who kept challenging me throughout this work and from whom I have learned so very much in the process.

Furthermore, I would like to thank all the people in the group of Robert Konrat for providing a welcoming and familial work environment and the support I have received from each and every one of them. Finally, I would like to thank Dariush Hinderberger and his team at Universität Halle for their contributions through our collaboration.

1 Abstract

While proteins have long been assumed to require a stable three dimensional structure to function a recently emerging class referred to as IDPs (intrinsically disordered proteins) has been shown to make due without it. Understanding their method of function is of great interest, because they are now known to be key players in important functional areas such as cellular signalling and regulation and can be linked to a multitude of human diseases. We have studied osteopontin (OPN), an IDP relevant to the development and progression of cancer, eliciting immune responses, bone remodelling and involved in a variety of signalling pathways.

Due to their intrinsic flexibility IDPs pose a challenge to many existing experimental approaches. NMR (nuclear magnetic resonance) has proven to be one of the most successful techniques in their study. Here we have developed a new method utilizing PRE (paramagnetic relaxation enhancement) effects in NMR. PRE provides information on long-range interactions in proteins. Utilizing a novel electron nucleus dipole-dipole interference term, which we refer to as $\Delta^1\mathrm{H}^\mathrm{N} - \Gamma_2$ or simply PRI (paramagnetic relaxation interference), we are able to detect concerted movement and global structural information on folded states in IDPs. Using this technique we find that OPN contains cooperatively folded compacted states, which encompass a central region as well as either the N-terminal or C-terminal region, but not both at once. Furthermore we were able to determine that the C-terminal region is less compact than the rest of the protein. This novel methodology allows to gather new insight into the structure of proteins in general and in particular can be used to enhance our understanding of IDPs.

Contents

2 Introduction

2.1 Intrinsically Disordered Proteins

The term IDP (intrinsically disordered protein) refers to a large subset of known proteins, which are fully or partially unfolded under native conditions as well as in their functional state. As their existence demanded a re-evaluation of the structure-function paradigm, which has long been considered an irrevocable foundational concept in biology, this class of proteins has long been neglected. [1]

Around the dawn of the 21st century they have become a more widespread research topic. Evidence has been accumulating ever since, which suggests that rather than being unstructured in the sense of behaving like a random coil these proteins exist in ensembles of rapidly interconverting and highly dynamic structural states. [2] It should be noted that protein structure can best be described by a structural continuum, particularly in case of IDPs, as even compacted states observed in them typically are still fairly flexible. [3] Therefore a more adequate determining criterion of IDPs is not the lack of rigid secondary or tertiary structure but rather their dynamics. [4]

Both the questions what might cause proteins to be unfolded under native conditions as well as how they might work, considering the framework of the structure-function paradigm, have been of great interest to the field of structural biology in recent years. Contrary to our traditional understanding of the importance of structure it is the current belief that their intrinsic flexibility actually makes IDPs particularly suitable for certain functions. One prominent example of this phenomenon is ligand binding. While this is not a task IDPs are frequently associated with it is one they excel at. IDPs tend to be promiscuous binders, i.e. they are able to recognize a variety of ligands without loss of specificity.

A majority of IDPs are believed to be involved in ligand binding, transcriptional and translational regulatory events as well as cell signalling, while enzymatic catalysis is usually reserved to well-structured proteins. IDPs are more prevalent in higher complexity organisms. Uversky et al. pointed out that the increased need for the functionality commonly associated with them might be the reason for that. [3, 4]

Multiple clues towards the reason for the unusually dynamic behavior of this fairly large subset of known proteins have been discovered. It has been shown that IDPs have a characteristically different amino acid composition than stably folded proteins. In particular they contain a significantly higher percentage of charged amino acids combined with much lower overall sequence complexity, as quantified by Shannon entropy. [5] In fact, the differences in the primary sequence are so striking they can even be used to predict whether a protein is natively folded or not. [6]

The experimental characterization of IDPs is challenging due to their highly dynamic behavior. While x-ray crystallography is undoubtedly the domineering technique for protein structure determination it is not suitable for the study of IDPs as it requires highly organized crystals, resulting in dynamic regions only being recognized by a lack of electron density. Therefore the only information one can gather on flexible regions, and by extension also IDPs, by this technique is the knowledge of their existence. [3] As many IDPs adopt a more rigid conformation upon ligand binding one commonly applied workaround is to study bound states rather than the flexible ensemble. This is somewhat unsatisfactory, for one because it does not work for all IDPs and for another because it only manages to capture a glimpse of a specific situation but fails as an image of the characteristic flexibility that is defining for IDPs. Therefore researchers depart to alternative techniques in the study of those proteins.

Many of the experiments applied in the early days of IDP research have originally been developed for the characterization of denatured proteins in the study of the protein folding process. The reason this has been fairly successful is that both sets of proteins share the lack of a stable 3D structure as a defining characteristic. Structural studies on either one of those sets of proteins typically aim to characterize them through properties of the ensemble. [7] Among a wide array of biophysical methods applied to characterize IDPs [7] one of the most frequently and successfully used ones is biomolecular NMR. [8] The NMR parameters used in the structural study of IDPs, such as nuclear Overhauser effect (NOE), residual dipolar couplings (RDC) and relaxation rate constants, are also commonly used in the structure determination of well-folded proteins. [9]

One problem all experimental techniques share in the study of IDPs is that

the measurements reflect a time as well as ensemble average. The time average is a result of internal motions in the observed molecules in the duration of the experiments. The ensemble average results from the differences between copies of the proteins in the sample, as at least to date limitations in sensitivity prevent single molecule structural studies. Especially when considering the high degree of flexibility associated with IDPs it becomes apparent that these averaging effects run at risk of covering up interesting features of the studied proteins. It is however not only experimental methods, which find themselves challenged by IDPs, but also simulations face hurdles, even though they are time resolved and do not face the problem of ensemble averaging.

Facing experimental limitations in the study of natively folded proteins researchers expanded their methods by a theoretical approach known as molecular dynamics simulations. It has been used successfully as a complementary method, providing valuable insight into areas inaccessible to experiments. [10, 11] Their single molecule resolution makes simulations promising in the study of IDPs as well because it circumvents the limitations due to averaging effects found in experiments, however the extent of their conformational space does require elaborate sampling techniques pushing the limits on current computational capabilities. Another important challenge in the study of IDPs via de novo molecular dynamics simulations lies in the accuracy of force fields. Currently used force fields have been optimized for the simulation of stably folded proteins and are known to overstabilize helical structures as well as peptide-peptide interactions, an issue that is much more pressing when working with IDPs. [12]

2.2 Osteopontin

OPN belongs to the SIBLING (Small Integrin-Binding LIgand, N-linked Glycoproteine) family of proteins. The family is believed to be the result of early gene duplication followed by divergence, a fact that is hinted at by the term SIBLING. Currently five members of the family are known, specifically osteopontin (OPN), bone sialoprotein (BSP), dentin matrix protein I (DMPI), dentin sialophosphoprotein (DSPP) and matrix extracellular phosphoglycoprotein (MEPE). On a primary sequence level they are poorly conserved, yet a number of sequence motifs as well

as a tendency towards acidic amino acids remain present across species in all SIBLING proteins. All SIBLINGs are small, soluble, RGD-motif containing secreted proteins. They act as promiscuous binders, being ligands for integrin receptors (both via RGD and non-RGD interaction), matrix metalloproteases (MMPs) [13] and complement factor H. [14, 15]

Sequence similarity is not only comparably small between different SIBLINGs but at least in case of OPN also the orthologues lack significant similarity on this level. Additionally to the features linking all SIBLINGs special conserved characteristics of OPN include enrichment of aspartic acid [15] as well as a conserved thrombin cleavage site (RSK) located approximately in the middle of the sequence. Cleavage uncovers the RGD motif in the N-terminal fragment thus aiding the interaction of OPN with integrins. OPN is able to bind a number of integrin receptors, including $\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_4\beta_1$, $\alpha_5\beta_1$, $\alpha_8\beta_1$, $\alpha_9\beta_1$. [14] The C-terminal fragment produced by thrombin cleavage contains a binding motif for the receptor CD44. [16]

Very much fitting its first description by Senger in 1979, which characterized it as a secreted phosphoprotein in malignant mammalian epithelial cells, [17] OPN received significant attention for its involvement in cancer. [14, 16, 18, 19] The SIBLING family in general is linked to various types of cancer, particularly to metastasis formation and progression. [14, 15] OPN is known today to be involved in virtually all steps of tumor progression, however the underlying molecular mechanisms are largely unknown. [14, 20] In healthy cells it was originally suspected to be merely responsible for mineralization of teeth and bones, however OPN is now known to be a key player in cellular signalling. Many of its important regulatory functions are achieved by acting as a ligand to ubiquitous cell surface receptors such as the above mentioned integrin receptors and CD44, which among other things mediate cell adhesion, migration and survival. [18]

OPN also carries out important functions in the immune system. The first role in immunity it was associated with was that of a cytokine produced by activated T lymphocytes. It is expressed by a multitude of cell types in the immune system, including B and T lymphocytes, macrophages, neutrophils, dendritic cells and NK cells. In response to inflammation and injury its expression is upregulated in a large number of organs.

OPN is strongly upregulated in macrophages as well as in their surroundings and has been shown to have a diverse role in regulating macrophage function with respect to their survival, migration, activation, phagocytosis, as well as production of pro-inflammatory cytokines. [21,22] It is an important early chemoattractant in the recruitment of macrophages to sites of injury or infection, thus supporting the inflammatory response. Once macrophages reach their destination OPN becomes relevant for their activation. [21, 23] Via an interaction with $\alpha_v\beta_3$ integrin OPN stimulates macrophages to secrete the pro-inflammatory cytokine IL-12, while suppressing the production of IL-10 through interaction with CD44. This change in cytokine expression correlates with an increase of T_H1 immunity. [18, 24]

Another cell type where OPN is highly expressed are immature dendritic cells, which function as a first line of defense in innate immunity. Mature dendritic cells show significantly reduced OPN expression levels, but OPN is suspected to be involved in inducing the maturation process. [21, 23] It is involved in homing dendritic cells to draining lymph nodes and supports the production of two major regulators of cell-mediated immunity, IL-12 and TNF-α. [21, 24, 25] In the interaction of dendritic cells and T cells OPN causes a bias towards T_H1 over T_H2 polarization of naive T cells through elevated levels of IL-12 and TNF-α and increased IFN-γ production by T cells. [25]

OPN is present at high concentrations in mineralized tissues, including teeth and bone. It facilitates adhesion of osteoblasts and osteoclasts to the mineralized bone extracellular matrix. While it is present at high levels in developing osteoblasts and osteoclasts and mediates substrate attachment in them its absence does not cause malfunctions in the development of mineralized tissues. [23] OPN is known to be an important regulator in osteoclast mediated bone resorption. Through interaction with $\alpha_v\beta_3$ integrins it stimulates demineralization. [23, 26] The ability of OPN to promote demineralization can lead to loss of bone mass and related illnesses such as osteoporosis, but it is also a potent effector in reducing or preventing ectopic vascular calcification. [18, 26]

NMR experiments of an N-terminally truncated version of OPN under in vitro conditions have revealed that it lacks a stable three dimensional solution structure and it has been proposed that this may be intricate to its diverse role. [15] Today it is classified as an intrinsically disordered protein (IDP). While it is highly

flexible and shows no single stable tertiary structure under native conditions it is well established that at least this version of OPN samples extended as well as partially structured and significantly compacted states. [27] These states are characterized by local secondary structure elements of significantly reduced flexibility and possess a compact core, however the organisation of the folded states is mostly unknown. [27, 28]

2.3 Paramagnetic Relaxation Enhancement

In pulsed NMR experiments nuclear magnetic moments are oriented in accordance with a strong static magnetic field (B_0 field) in an equilibrium state until a RF (radio frequency) pulse, a short lived magnetic field which points in another direction, disturbs this state. The processes leading to the system returning to its equilibrium orientation are termed relaxation. As we detect only components of the magnetization transverse to the B_0 field the signal declines as the system moves back to its equilibrium state with spins in longitudinal orientation. The presence of paramagnetic species, i.e. unpaired electrons, in the sample increases the relaxation rate of nearby nuclei. This phenomenon is often observed as a nuisance, for instance as overall signal reduction due to dissolved oxygen, but it can also be used to obtain important structural information.

There are three phenomena that can be caused by the paramagnetic nature of a sample, namely paramagnetic relaxation enhancement (PRE), residual dipolar couplings (RDC) and pseudo contact shifts (PCS). Among those effects PRE and PCS are limited to paramagnetic systems, while PDC can also appear in diamagnetic systems. PRE occurs in all paramagnetic systems while RDC and PCS can only be caused by paramagnetism associated with anisotropic electron g-tensors. [29] An electron has an anisotropic g-tensor if it possesses both spin and orbital angular momentum. [30] Examples for isotropic systems are nitroxide spin radicals and EDTA-Mn^{2+} whereas many metal ions are anisotropic systems. [29]

Particularly when working with paramagnetic labels in biomolecules isotropic systems provide a number of practical advantages, the first of which being that due to the absence of PCS no separate assignment of the paramagnetic sample is required. This is not only an advantage because generating reliable assignments

is a laborious task but also because due to the signal reduction in paramagnetic samples it can actually prove very challenging. Furthermore, the analysis of relaxation effects in isotropic systems is simplified by the the fact that Curie-type spin relaxation is negligible. [29,31] In experiments like this PRE is the only effect based on paramagnetic species which remains. The term paramagnetic relaxation enhancement (PRE) describes the phenomenon that the presence of paramagnetic species in the sample increases the rate of relaxation of spins in its proximity. This effect is detected as a loss of signal compared to a diamagnetic reference. While strong signals are generally desired when using PRE one aims to actively increase the rate of this relaxation at the cost of loss of intensity to obtain valuable information on spatial arrangements. The phenomenon of paramagnetic relaxation was first described by Solomon. [32] It is typically described using the Solomon-Bloembergen equations

$$\Gamma_1 = \frac{2}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 g^2 \mu_B^2 S(S+1) J_{SB}(\omega_I)$$

$$\Gamma_2 = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 g^2 \mu_B^2 S(S+1) \left\{4J_{SB}(0) + 3J_{SB}(\omega_I)\right\}$$

where μ_0 is the permeability of vacuum, γ_I is the gyromagnetic ratio of the nuclear spin, g is the electron g-factor, μ_B is the electron Bohr magneton and s is the electron spin quantum number. The function $J_{SB}(\omega)$ is the reduced spectral density function, defined as

$$J_{SB} = r^{-6} \frac{\tau_c}{1 + (\omega\tau_c)^2}$$

with τ_c, the correlation time, being defined as

$$\tau_c = \left(\tau_r^{-1} + \tau_s^{-1}\right)^{-1}$$

The variable τ_r refers to the molecule's rotational correlation time and τ_s the ef-

fective relaxation time of the electron. [29] It is the transverse relaxation rate Γ_2 that is typically detected in PRE studies, as it is of higher intensity than the longitudinal relaxation rate Γ_1. It is directly measured as

$$\Gamma_2 = R_{2,\text{para}} - R_{2,\text{dia}}$$

where $R_{2,\text{para}}$ is the transverse relaxation rate in the paramagnetic sample and $R_{2,\text{dia}}$ the corresponding diamagnetic reference. Quite frequently the intensity ratio (peak height) between paramagnetic sample and diamagnetic reference is also referred to as PRE although this is not precisely correct and cannot be used for quantitative comparison of rates. [29]

The primary mechanism driving PRE is dipole-dipole relaxation. While the dipole-dipole interaction of electron and nucleus dramatically increases the relaxation rate of the nucleus. Its effect is negligible for the electron as it is proportional to the magnetic moment of the spins, which is approximately three orders of magnitude larger in electrons than in nuclei. Data from PRE measurements is relevant to structural studies of macromolecules as it provides long-distance information in the range of 10 Å to 35 Å - depending on the paramagnetic species - which is not accessible to other common NMR experiments such as the nuclear Overhauser effect (NOE). [33]

As many biological macromolecules typically do not contain paramagnetic species, that is unpaired electrons, themselves in order to perform PRE measurements paramagnetic labels are attached to the macromolecules. Unpaired electrons are found in metal ions as well as organic radicals, both of which can be used to study proteins. There exists a variety of paramagnetic labels, which can be attached to the proteins in specific locations. The most common type of paramagnetic label is linked to cysteine residues using disulfide bonds. [29, 34]

2.4 Purpose of this Work

The purpose of this work was dual. For one, the goal was to analyse the intrinsically disordered protein OPN in the hopes of gaining further insight into its dynamic solution structure. For another, this project was concerned with param-

agnetic relaxation enhancement as a technique. Our goal was to analyse the effect the introduction of an additional paramagnetic label into the protein would have on the observed PRE, striving to understand how the combination of two labels affects relaxation fundamentally different than single labels. We aimed to develop a new experimental approach utilizing this information, which we could then use to further the knowledge of the first topic, the solution structure of OPN. This goal was pursued through a time efficient series of two-dimensional measurements leading to approximate data, as well as a more costly direct measurement of the relaxation effects. Our results indicate the importance of direct measurement and provide a protocol for the application of the method in future research.

3 Materials

3.1 Media for Bacterial Growth

LB	20 g LB-Medium Dissolve in 1 L H_2O Autoclave for 15 min at 121 °C
M9 10x	67.8 g Na_2HPO_4 x 2 H_2O 30 g KH_2PO_4 5 g NaCl Dissolve in 900 mL H_2O Adjust to pH 7.4 Add H_2O to a final volume of 1 L Autoclave for 15 min at 121 °C
M9	1:10 dilution of M9 10x (total volume 800 ml) Prior to use add: 10 mL 20 % Glucose 10 mL Trace Elements 2 mL $MgSO_4$ (1 M) 100 µL $CaCl_2$ (1 M) 800 µL Antibiotics 0.8 g $^{15}NH_4Cl$
LB-Agar plates	40 g LB-Agar Dissolve in 1 L H_2O Autoclave for 15 min at 121 °C

3.2 Buffers and Solutions

Agarose Gel	0.3 g Agarose 30 ml TBE 1x 0.5 µL SYBR Safe (Thermo Fisher Scientific) Mix agarose into TBE. Heat until agarose is entirely dissolved. Cool solution down to 60 °C Add SYBR Safe, mix well. Polymerize for 20 min, protecting it from light.
Ampicillin	1 g ampicillin add H_2O to a final volume of 10 mL Filtration (0.2 µm) Store at −20 °C
Ascorbic Acid 0.5 M stock	88.1 g ascorbic acid dissolve in 900 µL H_2O Adjust to pH 6.5 Add H_2O to a total volume of 1 mL Store at 4 °C
$CaCl_2$ 1 M	11.1 g $CaCl_2$ Dissolve in 100 mL H_2O Autoclave for 15 min at 121 °C
Coomassie Staining solution	5 g Serva Blue G-250 500 mL Methanol 100 mL Acetic acid 400 mL H_2O
Destaining solution	30 % Ethanol 10 % Acetic acid 60 % H_2O

dNTP mix 2 mM stock	2 mM dATP 10 µL dATP 100 mM 2 mM dCTP 10 µl dCTP 100 mM 2 mM dGTP 10 µL dGTP 100 mM 2 mM dTTP 10 µL dTTP 100 mM
DTNB 20 mM	8 mg DTNB 1 mL Ethanol
Laemmli buffer	20 g SDS 60 g Tris x HCl 286.8 g Glycin Dissolve in 2 L H_2O
$MgSO_4$ 1 M	12.04 g $MgSO_4$ Dissolve in 100 mL H_2O Autoclave for 15 min at 121 °C
MTSL 50 mM	10 mg MTSL 764 µL DMSO
MTSL reaction buffer	100 mM sodium phosphate 1 mM EDTA pH 8.0
PBS 10x	20 mM KH_2PO_4 80 mM Na_2HPO_4 25 mM KCl 1.4 M NaCl 50 mM EDTA Adjust pH to 6.8 using concentrated HCl Autoclave for 15 min at 121 °C
PBS 1x	100 mL PBS 10x 900 mL H_2O If necessary adjust pH to 7.26

PBS 1x (pH6.5)	100 mL PBS 10x 900 mL H_2O Adjust pH to 6.5 using concentrated HCl
PBS High Salt	1 L PBS 1x 58.44 g NaCl
Protein Sample Buffer 2x	120 mM Tris x HCl (pH 6.8) 6 % (w/v) SDS 20 % (v/v) Glycerine 0.01 % (w/v) Bromphenol blue 10 % (v/v) β-Mercaptoethanol store at 4 °C
Separating gel SDS-PAGE	8.7 mL Tris HCl pH=7.0 26.0 mL 40 % PAA (37.5:1) 34.7 mL H_2O 346.7 µL 20 % SDS 50 µL TEMED 216.7 µL APS
Stacking gel SDS-PAGE	3.8 mL Tris HCl pH=6.8 3.8 mL 40 % PAA (19:1) 22.7 mL H_2O 75.8 µL 20 % SDS 50 µL TEMED 151.6 µL APS
TBE 10x	500 mM Tris 1.5 M NaCl Adjust pH to 7.5
TBE 1x	100 mL 10x TBE 900 mL H_2O
Trace Elements	5 g EDTA 0.83 g $FeCl_3x2H_2O$

84 mg $ZnCl_2$
13 mg $CuCl_2$x2H_2O
10 mg $CoCl_2$x2H_2O
10 mg H_3BO_3
1.6 mg $MnCl_2$x4H_2O
Dissolve in H_2O to reach a volume of 1 L

3.3 Protein

All experiments were performed on a 220 amino acid long metastasis associated N-terminal truncation mutant of quail osteopontin containing positions 47-264 of the native protein as well as a leading methionine. In the following this truncated protein is referred to as OPN220. Additionally to the wildtype sequence cysteine mutants were engineered by means of PCR. Point mutations were introduced at positions 54, 108, 188 and 247 where the numbers refer to the numbering in the full length protein. Single mutants as well as double mutants were created. The previously described vector pET11d-OPN220 was used in the creation of the mutants. [35] The accuracy of the sequences was asserted via sequencing.

4 Methods

4.1 Preparation of Mutant DNA

Mutations were introduced into OPN220 in the vector pET11d-OPN220 using PCR (polymerase chain reaction). All primers were ordered from Sigma Aldrich. Agarose gel electrophoresis was performed for quality control. To prepare agarose gels agarose was mixed with the buffer TBE and slowly heated until fully dissolved. The solution was cooled down to approximately 60 °C before 0.5 µl SYBR Safe (Thermo Fisher Scientific) dye was added. Gels polymerized for 20 min while being protected from light. Electrophoresis was performed at 100 V for 30 min and correct size of the PCR product was evaluated under UV-light. After agarose gel electrophoresis 1 µl of the restriction enzyme DPNI were added and incubated at 37 °C for 15 min to destroy the template.

DNA purification was performed with the PCR purification kit QIAquick (QIAGEN). The purified product was transformed into *E.coli* DH5α and grown over night at 37 °C on ampicillin containing plates. Colonies selected from the plates were grown over night at 37 °C in 20 ml LB medium before isolating the plasmids using QIAprep Spin Miniprep Kit (QIAGEN). The success of the procedure was evaluated by sequencing. Purified plasmids were stored at −20 °C.

4.2 Transformation of Competent Bacteria

For production of protein *E.coli* BL21(DE3) were used. Bacteria from stock (stored at −80 °C) were thawed on ice and 1 µl of approximately 100 µM vector DNA were added. Bacteria were kept on ice for 10 min, then heated to 42 °C for 45 s before being cooled on ice for another 2 min to conclude the transformation. 500 µl LB medium were added and the bacteria were grown for 1 h at 37 °C at 150 rpm. 100 µl of the cultures were then plated on ampicillin containing LB-Agar petri dishes and incubated at 37 °C over night.

4.3 Expression and Purification of ^{15}N-labeled OPN

A single transformed culture was picked from the plate and grown in 25 ml LB with ampicillin over night. 10 ml of this preculture were added to 800 ml LB and grown to an OD 600 of 0.8-1.0 at 37 °C, 150 rpm. Bacteria were harvested by centrifugation (4000 rpm, 20 min, 4 °C and the pellets then transferred to minimal medium M9, freshly prepared from M9 10x stock. Bacteria were grown in M9 at 37 °C and 150 rpm for 30 min before reducing the temperature to 27 °C. After 10 min protein expression was induced with 800 µl of 4 M IPTG and continued through over night culture (27 °C, 140 rpm).

Following harvesting by centrifugation (4000 rpm, 20 min, 4 °C) cells were resuspended in 20 ml PBS 1x, for cysteine mutants. DTT was added to a concentration of 1 mM. Cells were ruptured via sonication (2x 3 min, on:off = 1 s:1 s, 50 % amplitude) and fragments pelleted by centrifugation (18 000 rpm, 20 min, 4 °C). The supernatant was heated (10 min, 85 °C) and re-pelleted (18 000 rpm, 20 min, 4 °C). Cysteine mutants were treated with DTT again.

The supernatant was transferred and treated with 35 % ammonium sulfate for 30 min at 4 °C while slowly stirring the solution. After a subsequent centrifugation step (18 000 rpm, 20 min, 4 °C) the ammonium sulfate concentration in the resulting supernatant was increased to 50 % and the solution was again stirred slowly for 30 min at 4 °C. The pellet was collected in a final centrifugation step (18 000 rpm, 20 min, 4 °C) and quickly washed with ddH_2O, resuspended in 10 ml 1x PBS and DTT was added when working with cysteine mutants. To adjust the salt concentration dialysis against 2 L 1x PBS was performed over night at 4 °C.

Ion exchange chromatography was performed using a Resource Q (GE Healthcare) column. The sample was applied to the column and eluted using a gradient between PBS and PBS High Salt. The protein is eluted at approximately 25 % High Salt buffer. Purity of the protein in protein containing fractions was evaluated using SDS-PAGE. SDS gels consisting of separating gel and stacking gel were prepared beforehand and stored at 4 °C. Samples from protein containing fractions were mixed 1:1 with 2x SDS protein sample buffer (loading dye) and inserted into chambers in the gel. SDS-PAGE was performed using Laemmli buffer and applying a current of 200 V for 45 min. Finished gels were coated with

Coomassie staining solution for 10 min, then treated twice with destaining solution for 15 min to remove excess dye. Pure fractions identified from the gel were pooled and concentrated to a volume of approximately 2 ml using a Centricon (Amicon) with 3000 Da molecular weight cut-off (centrifugation at 4000 rpm, 4 °C). Protein concentration was established via NanoDrop (Thermo Scientific) by means of absorption spectroscopy, using the absorption at 280 nm.

4.4 MTSL Labeling

Cysteine-mutated proteins were tagged with the nitroxide label MTSL (see fig. 1). Purified protein samples were incubated with DTT at a 5-fold excess for 15 min at room temperature. After this incubation time proteins were separated from any free DTT using PD-10 desalting columns (GE Healthcare) with MTSL reaction buffer. Proteins were eluted in a volume of 3 ml buffer. Concentration of free thiols in the sample was measured by means of absorption spectroscopy at 412 nm. For this purpose 50 µl of the protein dissolved in MTSL buffer were measured against 50 µl pure MTSL buffer, both diluted in 450 µL DTNB measurement solution. That solution was composed of 985 µl MTSL buffer and 15 µl of the 20 mM DTNB stock solution. The concentration of free thiols was calculated from the measured absorption as follows:

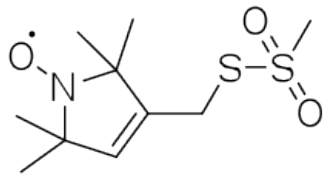

Figure 1: Chemical structure of the used paramagnetic nitroxide label MTSL

$$c = \frac{OD_{412}}{\epsilon_{OPN}} \times 10[\mu M]$$

Samples were treated with 2-fold excess of MTSL for a minimum of 3 h at 37 °C and 150 rpm. After this time the remaining concentration of free thiols was measured again, as described above. Absorption at this point was always ≤ 0.01 and thus deemed sufficient for our purposes. Protein samples were concentrated using Centricons (Amicon, 3000 Da molecular weight cut-off) and applying 4000 rpm at 4 °C until the volume was reduced to below 2 ml. This way the samples were

washed with NMR measurement buffer (PBS 1x, pH6.5) three times until finally being concentrated in that buffer. Final concentrations were measured by means of absorption spectroscopy, using the absorptive properties of the protein at 280 nm (calculated $\epsilon_{OPN} = 13\ 940$).

4.5 NMR Measurements

Measurements were performed on Varian spectrometers operating at 600 MHz and 800 MHz. Samples were measured at concentrations varying from 0.9 mM to 1.4 mM. OPN220 was measured in 1x PBS (pH 6.5) containing 10 % D_2O. Measurements were performed at room temperature (25 °C). All paramagnetic samples were measured in their paramagnetic state before reduction with ascorbic acid and then remeasured with identical measurement settings, albeit adjusted shimming and tuning. The reduction with ascorbic acid was performed with 5-fold excess of ascorbic acid over the spin labels (change in volume <5 %) for 1 h at 37 °C and 150 rpm. As a standard 2D measurement from which we derived PRE intensity data we used $^{1}H - ^{15}N - SOFAST - HMQC$ (band-Selective Optimized Flip-Angle Short-Transient heteronuclear multiple quantum coherence). To determine ^{1}H-T_2 relaxation we performed ^{1}H-T_2-HSQC with three relaxation delays (10, 30 and 50 ms) and interpolated the reduction of intensity with the exponential function $\frac{I}{I_0} = \exp(-R_2 t)$, as adapted from an approach by Clore and co-workers. [31] Spectra were processed with NMRPipe [36] and Sparky. [37]

4.6 Normalization of Intensity Based PRE Data

Peak heights observed in the paramagnetic samples (I_{ox}) were divided by heights observed in the same sample after reduction with ascorbic acid (I_{red}) giving intensity values as

$$I = \frac{I_{ox}}{I_{red}}$$

From here on all approximative data that was not altered any further will be referred to as not normalized. Due to aberrations of the resulting values from the

expected interval [0,1] additional normalization options were considered to allow for better comparability across samples and measurements. To allow for comparison of the PRE effect between samples the intensity ratios were then divided by a suitable factor to result in a distribution of values that was mainly focused on the interval [0,1] where it was assumed that the values that still disagreed with that interval were outliers. The factor by which values were assumed to be wrong was determined based on regions assumed to be unaffected by the PRE effect. Data resulting from this type of treatment will be referred to as interval normalized.

For measurements in the urea titration multiple approaches to normalization were considered. Additionally to not normalizing and interval normalizing the data normalization based on the total intensity in the measurement was considered. Normalization of intensities by division through the sum of all intensities will from now on be referred to as normalization by total intensity. Additionally, normalization by division through the sum of intensities of a region unaffected by the PRI was performed (normalization by constant region intensity).

4.7 Approximation of $^1HT_2$ Relaxation Rates

The (normalized) PRE data (see section 4.6) was used to approximate the components of the relaxation rates that occurred in the given samples due to the PRE effect by applying Newton's method. It was assumed that relaxation rates observed in paramagnetic samples are defined by two components as

$$R_{2,\text{para}} = R_{2,\text{dia}} + \Gamma_2$$

where $R_{2,\text{dia}}$ is the transverse relaxation rate without a paramagnetic label and $R_{2,para}$ is the transverse relaxation rate observed with the label. This leaves Γ_2 as the component of the transverse relaxation that occurs because of the paramagnetic labels. For the values of the $R_{2,\text{dia}}$ rate measured $^1HT_2$ relaxation rates of wildtype OPN220 were used. We applied Newton's method (first order, 10 000 iterations, initial guess of $\Gamma_2 = 4$) to calculate Γ_2 while indirectly determining $R_{2,\text{para}}$ from the peak intensities detected in HSQC measurements. Note that the chosen number of iterations is most certainly considerably higher than required but given the speed of Newton's method we did not feel the need to optimize it. It is pos-

sible to calculate Γ_2 in this way because of the indirect proportionality between intensity, as reflected by peak hight, and relaxation rate, i.e.

$$I_{ox} \approx \frac{1}{R_{2,para}}$$

$$I_{red} \approx \frac{1}{R_{2,dia}}$$

For the intensity ratios Wagner and co-workers [38] established that

$$\frac{I_{ox}}{I_{red}} = \frac{R_{2,dia} exp(-\Gamma_2 t)}{R_{2,dia} + \Gamma_2}$$

which we used as the key equation in the numerical approximation. Of the PRE data only those data points were included into the calculations that showed a normalized $I_{ox}/I_{red} > 0.05$ to circumvent numerical instabilities.

5 Results

5.1 PRE Effect

The results of the PRE measurements performed on OPN220 single mutants (see fig. 2) are widely consistent with those already published [27], despite differences in the calculation of the PRE. However, increased PRE effects distant from the labeling sites can be observed repeatedly, most pronounced in OPN220 C188, where the entire central region of the protein (120-160) shows virtually no signal in our measurements while being only diminished to half the intensity according to previous measurements. At the same time our measurements show residual signals in proximity of the label whereas previously published data did not. In contrast mutant C108 agrees with the previously published results fairly precisely. The remaining single cysteine mutants C54 and C247 show overall stronger PRE effect than was apparent from previous studies, however the effect is not nearly as prominent as with C188 and also more scattered along the protein chain. PRE values obtained from direct measurement of the rates (see fig. 3) show increased relaxation in the corresponding areas. To allow for better comparison of the intensity based PRE effect and the relaxation measurement we also determined the approximate increase of the relaxation rate using Newton's method (fig. 4). The obtained values are in good agreement qualitatively as well as in the same order of magnitude.

Measurements on all double mutants (see fig. 5) show diminished signal in the residues surrounding the labeling sites. Similar to the single mutants we also find PRE effects distant from the labeling sites, consistent with compaction in those regions. The regions affected in the double mutants are also mostly consistent with those found to be affected in the single mutants. Comparison of the PRE effect based on relaxation measurements and intensity based PRE effect (fig. 6) reveals the same features as observed in the single mutants.

To determine whether additional insight into the protein's structural properties could be obtained from the double mutants we used two approaches. First, in an approximative approach we extrapolated relaxation rates from the measured intensities similar to the approach described by Battiste and Wagner. [38] For

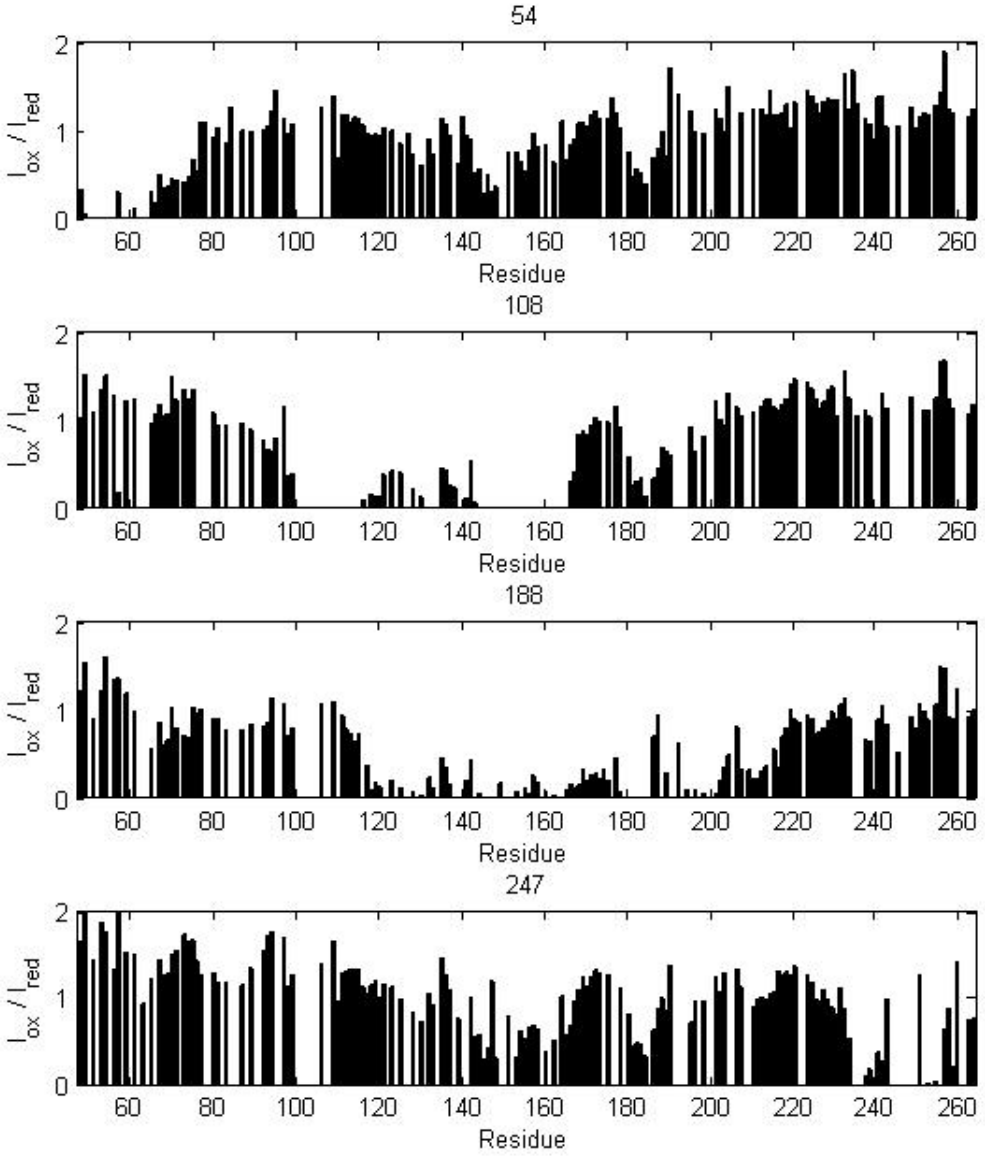

Figure 2: Approximative PRE effect as observed in the OPN220 single mutants C54, C108, C188 and C247.

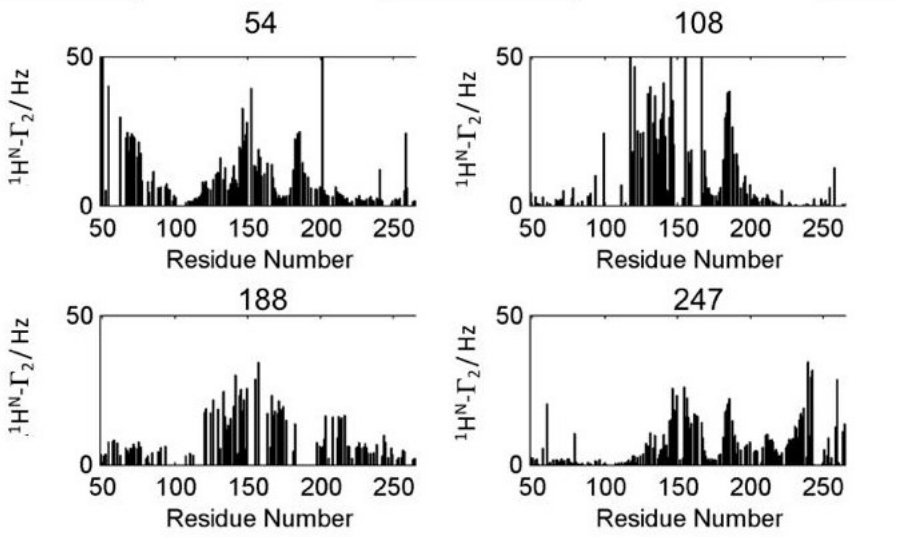

Figure 3: Directly measured PRE effect (relaxation measurement) as observed in the OPN220 single mutants C54, C108, C188 and C247. Reproduced from Ref. [39] with permission from the Royal Society of Chemistry.

details on the approximation process refer to section 4.7. These results (see fig. 5) suggest a component in the relaxation of the double mutants which cannot be explained by mere addition of a second label without some kind of interference effect, presumably cross-correlated cross-relaxation. The results of further analysis of this effect will be described in the next section.

5.2 $\Delta^1H^N - \Gamma_2$ - Paramagnetic Relaxation Enhancement Interference

Analysis of the PRI effect based on rate approximation (see fig. 7) reveal both positive as well as negative $\Delta^1H^N - \Gamma_2$. The most significant negative effect is observed in the mutant C54-C188 in the protein's central region. All variations of OPN220 with a mutation in position 247 show significant positive PRI but also some negative values. Using $^1H - T_2$ relaxation data measured in 1.4 mM samples of OPN220 C54, C247 and C54-C247 we determined that the approximative Γ_2 values were underestimated by a factor of 3.2 in those cases, presumably due to errors associated with normalization and approximation. Due to the fairly high potential error associated with the approximative $\Delta^1H^N - \Gamma_2$ values we supple-

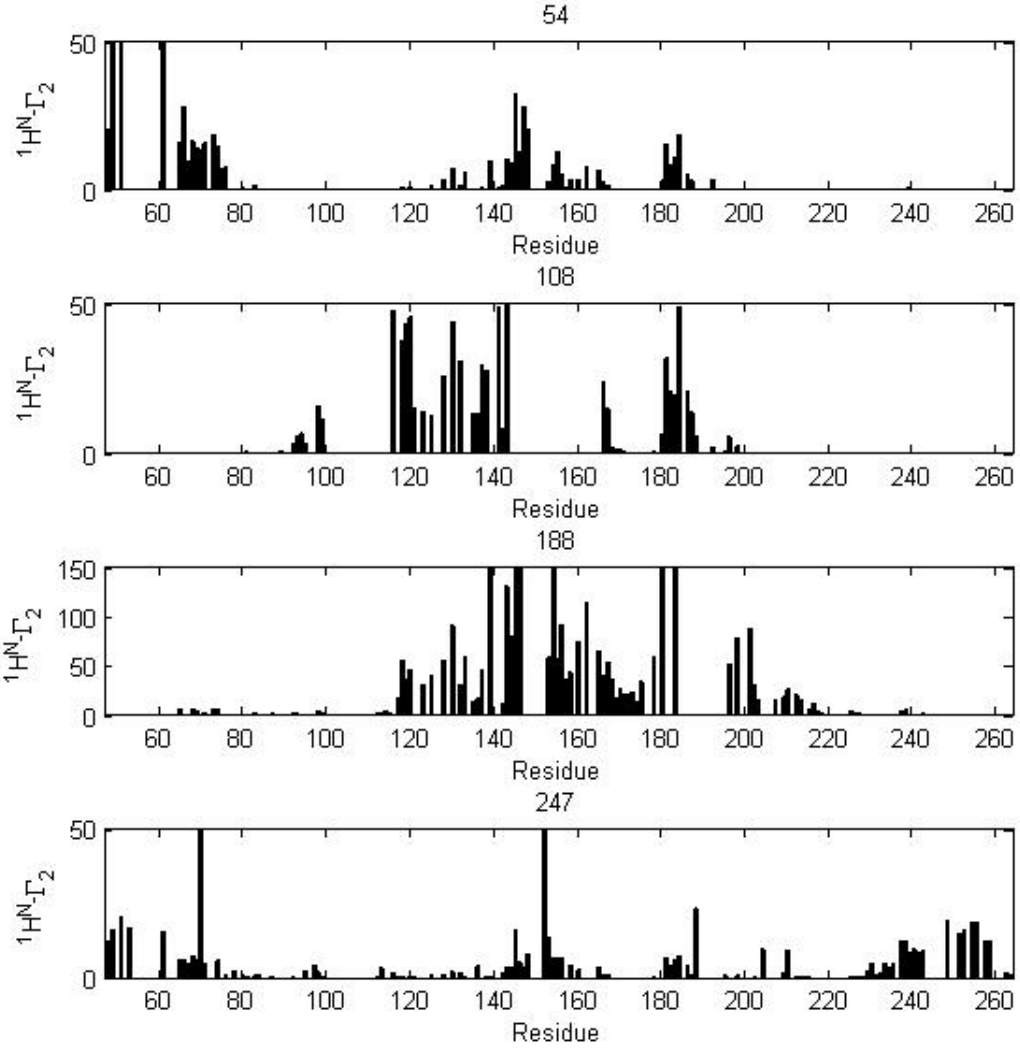

Figure 4: Approximative PRE effect as observed in the OPN220 single mutants C54, C108, C188 and C247. Rates are approximated based on intensity ratios between paramagnetic and diamagnetic state.

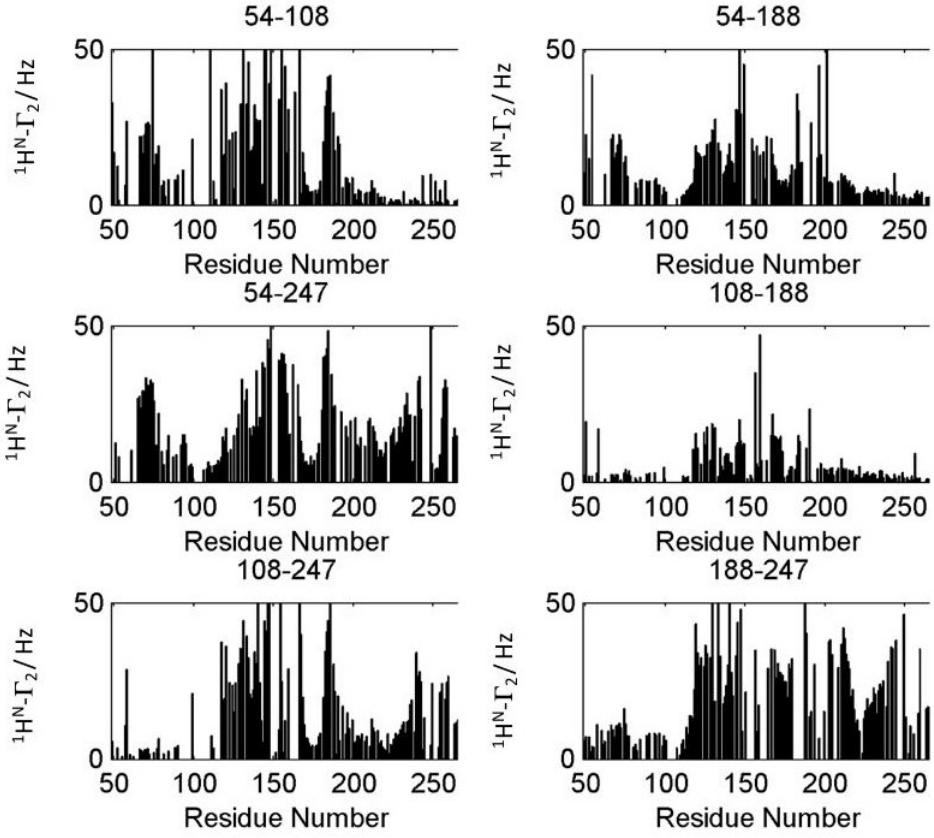

Figure 5: Directly measured PRE effect (relaxation measurement) as observed in the OPN double mutants C54-C108, C54-C188, C54-C247, C108-C188, C108-C247 and C188-C247. Reproduced from Ref. [39] with permission from the Royal Society of Chemistry.

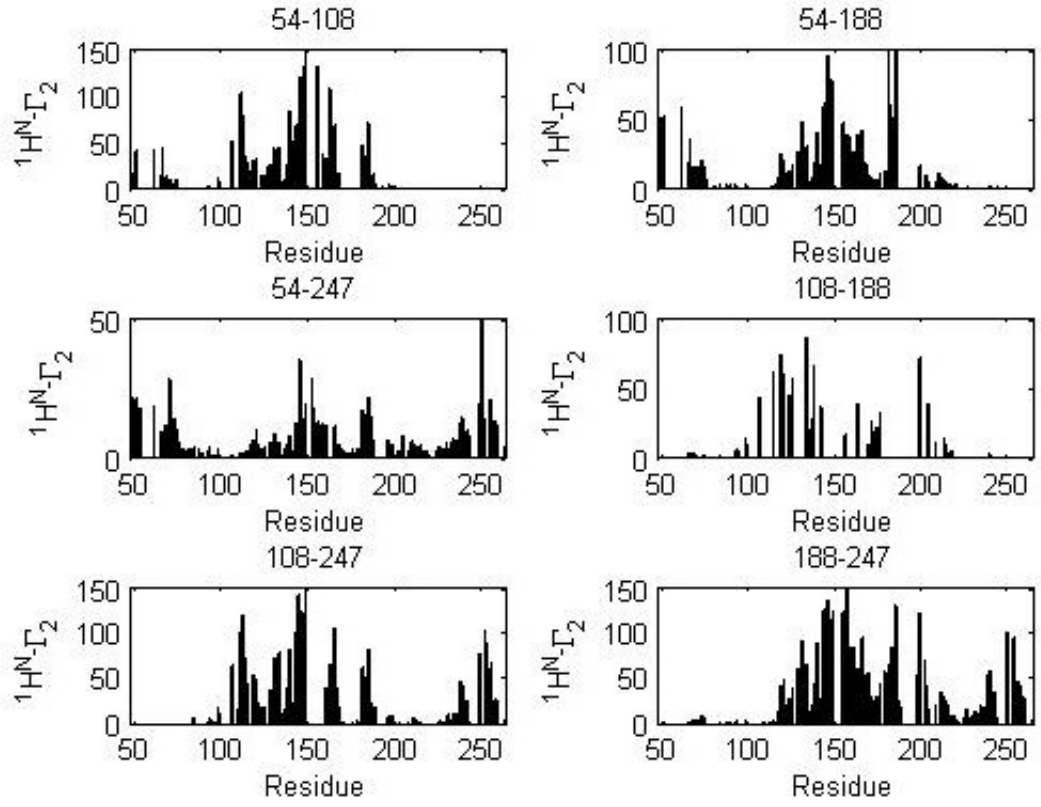

Figure 6: PRE effect as observed in the OPN220 double mutants C54-C108, C54-C188, C54-C247, C108-C188, C108-C247, C188-C247. Rates are approximated based on intensity ratios between paramagnetic and diamagnetic state.

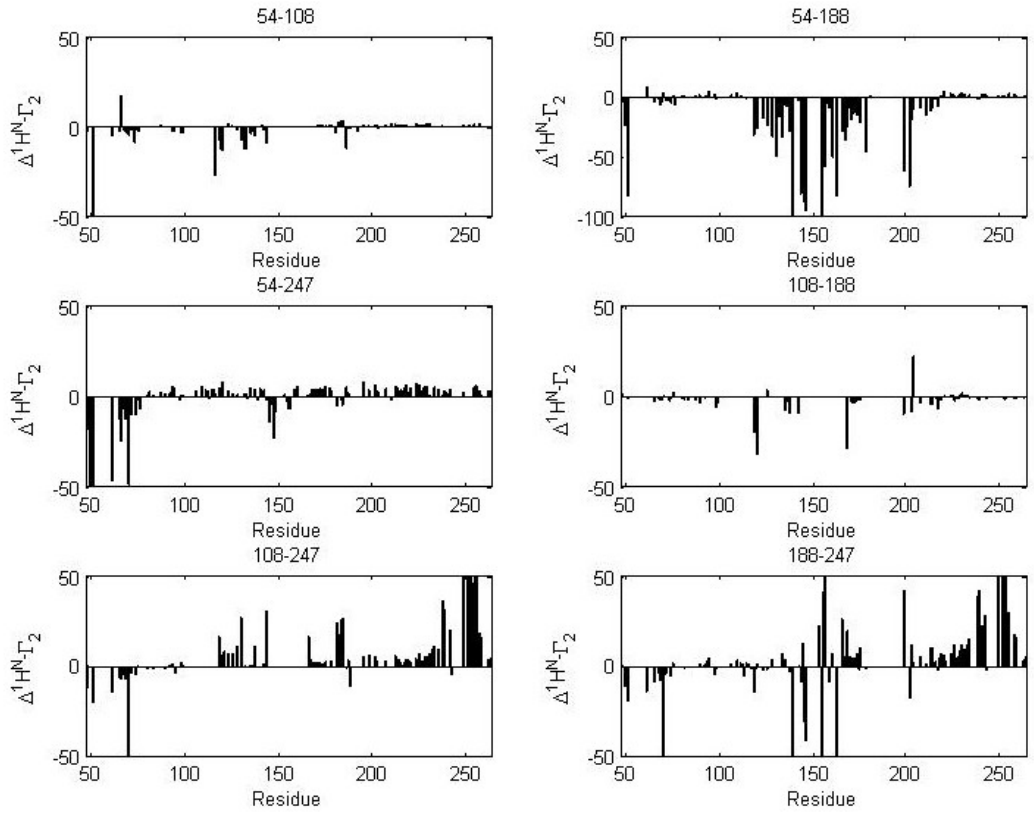

Figure 7: Approximative PRI effect of all six double mutants of OPN220. Rates were approximated based on intensity ratios between paramagnetic and diamagnetic state.

mented them with data based on T_2 measurements to get more detailed insight into the interference effect observed in the PRE measurements of the OPN220 double mutants. We performed T_2 measurements for all single and double mutants to obtain more quantitative data to describe the observed effect. The analysis of these measurements clearly shows a PRI$\neq$0 (see fig. 8). The OPN220 mutants C54-C247 and C188-C247 show PRI$>$0 while the others mostly show PRI$<$0. The most pronounced PRI effects were observed in the region of positions 120-180 while the effect appears diminished towards the termini of the protein, especially the C-terminus. Comparison of the $\Delta^1H^N - \Gamma_2$ effect based on approximation and direct measurement reveal significant differences of varying extent in all mutants.

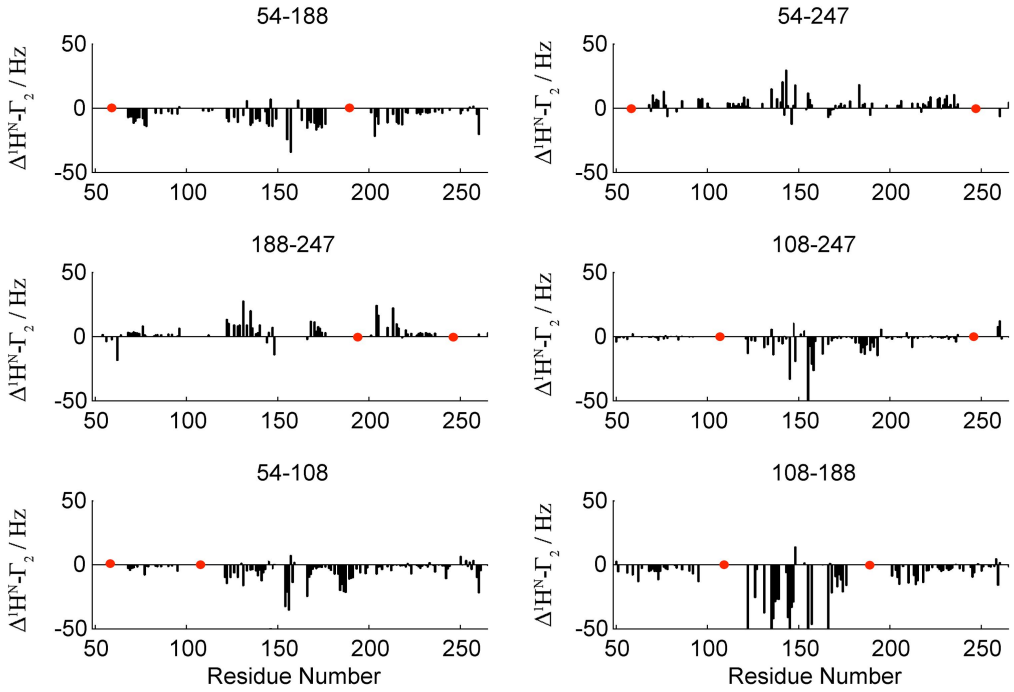

Figure 8: Visualization of the observed Δ^1H^N (PRI) of all six double mutants of OPN220. Red dots indicate labeling sites. Reproduced from Ref. [39] with permission from the Royal Society of Chemistry.

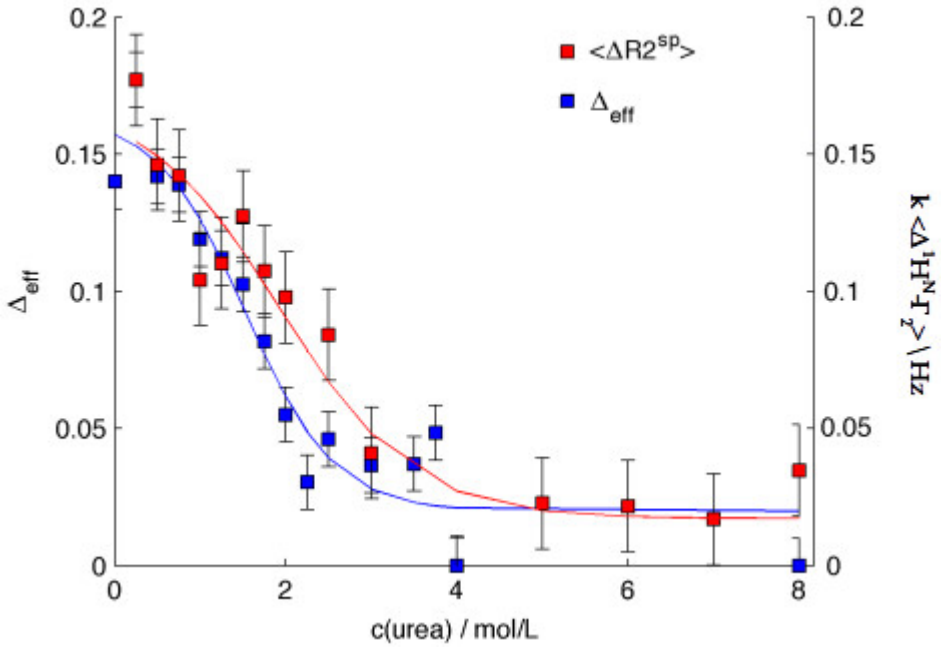

Figure 9: Denaturation of OPN220. Red data points reflect approximated PRI data, weighted with a proportionality factor $k = -\frac{1}{12}$ to allow for comparison of the two curves. Blue data points are obtained from DEER measurements. Δ_{eff} is the effective modulation amplitude of the DEER time trace. It is obtained at the maximum feasible dipolar evolution time, after experimental background correction. Lines are sigmoidal interpolations of the corresponding data points.

5.3 Urea Denaturation

The T_2 relaxation rates were approximated based on the normalized intensity ratio using Newton approximation (see section 4.7) for PRE data of OPN220 single mutants C54, C247 and the corresponding double mutant C54-C247. Due to time limitations no data from direct measurement of T_2 relaxation is available. The denaturation curve we observed with the approximated PRI shows good agreement with denaturation data previously obtained by means of EPR, using DEER experiments (see fig. 9). [27]

The data shows sigmoidal behavior with increasing urea concentration, indicative of cooperative unfolding. The fact that the denaturation curve based on the observed loss of PRI closely follows that obtained via EPR suggests that the PRI effect provides insight into cooperatively folded, i.e. compacted states. Fig. 10

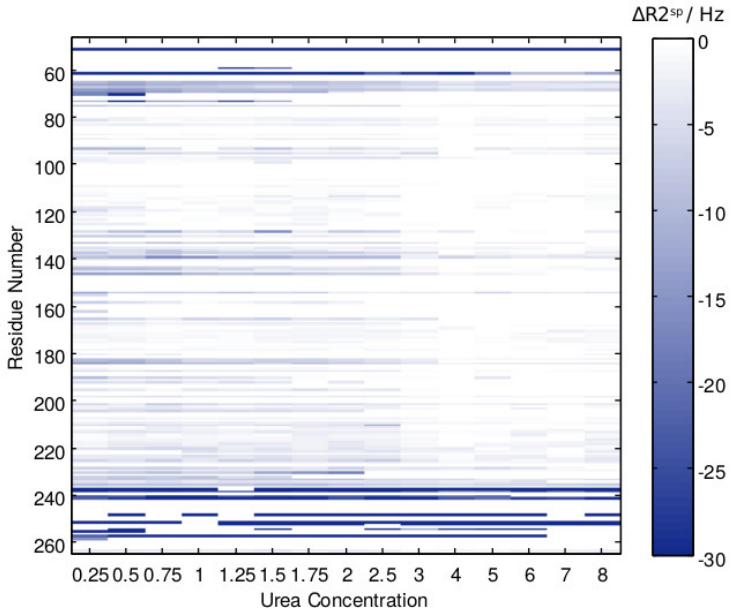

Figure 10: Analysis of approximated PRI effect in OPN220 C54-C247 double mutant with increasing concentration of urea.

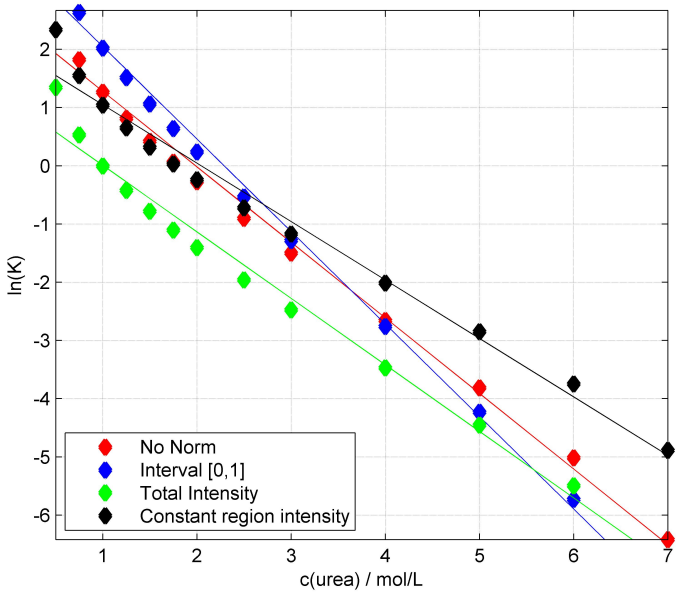

Figure 11: Evaluation of m-value in a urea titration of OPN220 mutant C54-C247 based on approximative PRI.

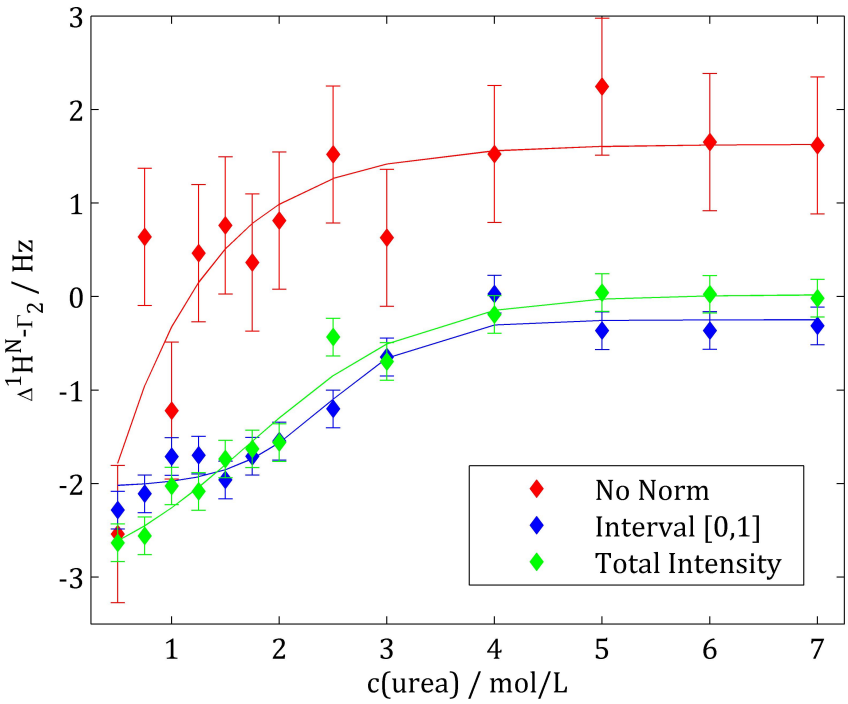

Figure 12: Evaluation of the urea titration of OPN220 mutant C54-C247 based on approximative PRI. Error bars indicate RMSE.

shows that the signal does not even fully disappear at 8 M urea. It should be noted that these approximation based results are by no means quantitative as is easily apparent from the difference between the values at the first measurement in the urea titration (0 M urea) compared to the measured PRI effect (see fig. 8). As a further measure for the influence of urea on the structure and stability of OPN220 we performed m-value analysis (see fig. 11 and tab. 1), finding a fairly consistent m-value throughout various normalization procedures. The normalization procedures do however result in fairly big differences between the titration curves (see fig. 12), most significantly in case of normalization by constant region intensity (data not shown).

Table 1: Results from the analysis of the approximative PRI effect observed in a urea titration of OPN220 mutant C54-C247.

Normalization	m-Value [L/mol]	ΔG_0 [J/mol]
none	-1.2973	-6379.86
Interval [0,1]	-1.5886	-9007.57
Total Intensity	-1.1423	-2846.20
Constant region intensity	-1.0033	-5080.22

6 Discussion

6.1 Relevance of this Work

The main application of biomolecular NMR is the study of the solution structure of proteins. One of the great strengths of the technique is that it enables the detection of transient structures as well as transitions between structures. It has been applied in the study of flexible proteins, including IDPs, with great success. So far all structural constraints obtained from NMR studies have been measures of distance or contact between two spins. Our approach suggests a new analysis which allows the detection of proximity with respect to three spins based on PRE. By introducing two paramagnetic labels we were able to determine correlated motion in the studied protein by exploiting a novel electron-nucleus dipole-dipole relaxation term. The basic idea is that in a situation where both paramagnetic centers are in proximity to a nuclear spin due to the mentioned electron-nucleus dipole-dipole interference term relaxation of the spin is affected. As outlined in fig. 13 on the example of a hypothetical IDP this allows us to probe whether a compacted state exists in the conformational ensemble where both paramagnetic labels are in the proximity of other residues at the same time. Traditional PRE measurements using only one paramagnetic center are unable to determine whether the observed relaxation results from an ensemble containing the hypothetical states I and II or a compacted state III exists.

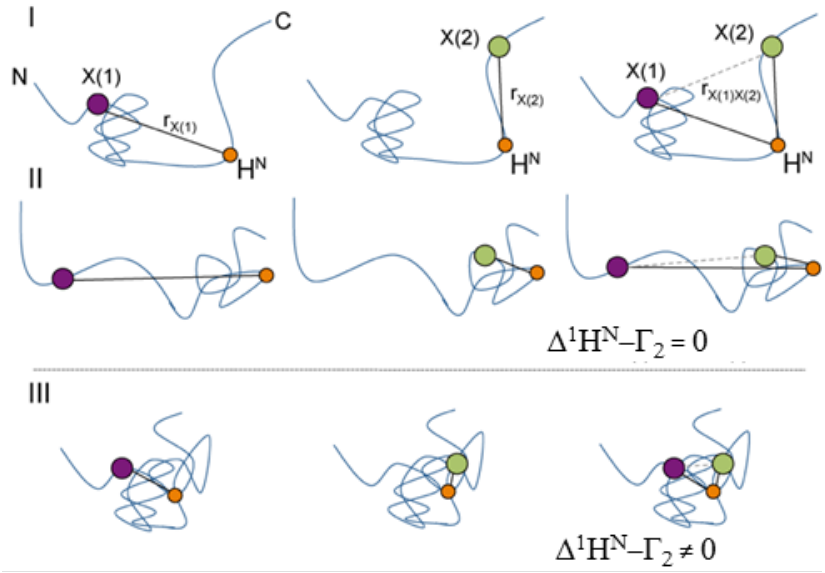

Figure 13: Schematic representation of an IDP in partially extended (I and II) and fully compacted (III) states. Paramagnetic spin labels are represented by purple and green dots and referred to as X(1) and X(2). An observed residue is highlighted by an orange dot. (I) reflects a situation in which neither spin labeling site is in spatial proximity to the amide of interest (H^N). (II) Shows one labeling site being close to the H^N but the other remaining distant. In (III) both spin labels are close to H^N simultaneously, thus resulting in a non-zero PRI. Reproduced from Ref. [39] with permission from the Royal Society of Chemistry.

6.2 Methodological Considerations

The use of PRE methods has been validated by the successful application to a number of problems, however in the application to IDPs it suffers from the intrinsic limitation to display an average over all existing structural substates. The new PRE-based method we suggest manages to reduce this problem by utilizing the interference effect between two paramagnetic labels to characterize the structural ensemble. The observed change in relaxation is the result of an electron-nucleus dipole-dipole interference term. The strong distance dependence of dipole-dipole relaxation with r_{HN}^{-6} makes the method selective for compacted states. The term we refer to as paramagnetic relaxation enhancement interference (PRI) or $\Delta^1H^N - \Gamma_2$ describes the deviation of the relaxation of a nuclear spin under the influence of two paramagnetic centers from the cumulative effect of the two paramagnetic centers, i.e.

$$\Delta^1H^N - \Gamma_2 = {}^1H^N - \Gamma_2[X(1) + X(2)] - \left\{{}^1H^N - \Gamma_2[X(1)] + {}^1H^N - \Gamma_2[X(2)]\right\}$$

where ${}^1H^N - \Gamma_2[X(1) + X(2)]$ is the transverse relaxation rate of an amide proton ${}^1H^N$ in a double mutant containing spin labels at residues X(1) and X(2) and ${}^1H^N - \Gamma_2[X(1)]$ and ${}^1H^N - \Gamma_2[X(2)]$ the transverse relaxation rate in the single mutants with spin labels at X(1) and X(2) respectively.

Suppose there were no compacted state in the ensemble where X(1) and X(2) are close to each other then there can be no cross-correlation between them, thus $\Delta^1H^N - \Gamma_2 = 0$. If $\Delta^1H^N - \Gamma_2 \neq 0$ is observed this suggests that there exists a conformation in which some ${}^1H^N$, X(1) and X(2) are in proximity of one another. The magnitude of the interference term $\Delta^1H^N - \Gamma_2$ is strongly dependent on $< 3\cos^2\theta - 1 >$ where θ describes the angle between the dipoles corresponding to $X(1) - {}^1H^N$ and $X(2) - {}^1H^N$ and the angles indicate the time and ensemble average. The knowledge of the proportionality $\Delta^1H^N - \Gamma_2 \propto < 3\cos^2\theta - 1 >$ allows for an interesting structural interpretation of the PRI effect. In very rigid systems it enables the detection of the angle θ and is fairly straight forward. However, as we are dealing with flexible molecules which show no single well-defined structure we need to analyse the averaging effect, which results from the molecules' motions.

The interpretation of the PRI in IDPs is complicated by the multitude of

averaging effects. To better understand the averaging of the angle θ between the interacting elements in the PRI we consider Flory's polymer theory. Based on this theory we expect values around $\theta = 180°$ extended segments which considering the mentioned $\Delta^1H^N - \Gamma_2 \propto < 3\cos^2\theta - 1 >$ leads to $\Delta^1H^N > 0$. For globular states Flory's theory suggests $\theta = 90°$, which leaves us with $\Delta^1H^N < 0$. Therefore even in highly flexible systems the sign of the PRI gives insight into their structural organisation.

6.3 Intensity Based Results

Using intensity based PRE values we have determined a relaxation component $\Delta^1H^N - \Gamma_2 \neq 0$ for all mutants of OPN220 (see fig. 7). In a comparison of the PRE observed in the approximate intensity based approach versus the more time consuming direct measurement of relaxation rates we have established that the approximation is in good agreement with the direct measurement for all single as well as double mutants (see fig. 4 and fig. 6). One weakness of the approximation of rates is that there are more residues missing. This happens because regions, which are strongly affected by the PRE, particularly those close to the labeling site, can often not be detected in this approach or lead to extremely low values which have to be excluded to avoid numerical instabilities in the approximation of relaxation rates. While these results look somewhat promising it shall not remain unmentioned that the data obtained from approximation of ${}^1H^N - \Gamma_2$ can only be interpreted in a semi-quantitative way due to the high potential error resulting from the combination of imprecise data and numerical approximation. They are most useful when trying to determine quickly whether more thorough investigation would be promising.

As Clore and co-workers have pointed out [29] the intensity ratio between paramagnetic and diamagnetic state is not physically meaningful and cannot be used to make quantitative statements about relaxation. This is in itself already a big problem when aiming to analyse the PRI effect since the analysis requires calculations based on effects across different mutants, namely a double mutant and its respective single mutants. Assuming that each of those data sets is afflicted with a significant error any calculations based on them will necessarily be imprecise,

as will therefore the approximative PRI effect. This can partially explain why comparison between measured versus approximated effects reveals much better results of the approximation in case of the PRE rather than the PRI.

The apparent PRI effect based on approximation (see fig. 7) shows significant deviations from that seen in direct measurement (see fig. 8) with no apparent system behind the extent of the errors. The mutant C54-C188 in particular shows exceptionally good agreement between the methods. On the other hand C54-C108 and C108-C188 show virtually no PRI effect in the approximation but significant negative PRI in the direct measurement. We presume this to be an effect resulting from the above mentioned loss of data points due to the approximation procedure as well as numerical errors. Another interesting yet potentially problematic effect that appears in the approximation is the prediction of a strong PRI effect in proximity of the labeling site, as can be observed in the N-terminus of mutant C54-C247 and the C-termini of C108-C247 and C188-C247. These artefacts presumably result from the limitations in obtaining (reliable) data in regions strongly afflicted by the PRE effect. Knowing about this issue it might be reasonable to consider removing all strong PRI values observed in the proximity of paramagnetic labels when working with the approximation. If we were to do so both the results for mutants C54-C247 as well as C188-C247 could be considered fairly high quality approximations, however doing so brings the risk of removing important data and should therefore be considered cautiously.

Another issue we observed when working with intensity based PRE, the foundation of the approximative method, is that the intensity ratios do not deliver PRE values in the interval [0,1] as we would expect given that all relaxation mechanisms found in the diamagnetic sample are also present in the paramagnetic sample and that the only expected difference is additional relaxation caused by the paramagnetic label. Deviations were found up to intensity ratios being distributed in the interval [0,2] rather than [0,1]. Differences in the intensities between paramagnetic and diamagnetic measurement can either be a result of the sample composition or the measurement process. Since paramagnetic and diamagnetic measurements were obtained from the same samples differences in sample composition seem unlikely. In order to turn a paramagnetic into a diamagnetic sample ascorbic acid was added (see section 4.5). The minute volume that was added makes dilution

effects highly unlikely, changes in the pH were avoided by adjusting the solution of ascorbic acid to the pH of the buffer measurements were performed in. The remaining possibility for differences in the sample is degradation of the protein, however given the typically very short time between measurements (1.5 h or less) makes this unlikely. Another clue that makes the changes to the sample unlikely as a cause of the observed error is that the observed error appears to be random in nature, whereas the addition of ascorbic acid and the related treatment was systematic. This leads us to believe that the cause of the deviation from the expected interval might lie in the measurement process. There are a number of potential sources of error (e.g. pulse imprecisions, temperature fluctuations, shimming, tuning) associated with the measurement, which could explain the deviation, however we are not able to pinpoint the exact cause.

6.4 Urea Denaturation

Using the approximative PRI effect we have shown that the IDP OPN220 shows cooperative unfolding under the influence of urea. This proves that while they cannot be characterized by a single stable structure IDPs are by no means fully extended or unfolded but do in fact contain cooperatively folded structures. Not only do these results back up the recent change in our understanding of the lack of a native structure in IDPs, but the fact that they are in good agreement with data obtained via DEER measurements also supports the validity of our new technique. Because of the mentioned issues with the interval based approach we tried a number of alternative measures to correct for this error.

To our surprise the differences between the normalization approaches were, at least according to the m-value analysis where we applied them, not exceptionally big. The titration curves however differed significantly, both in their absolute values as well as shape. The only normalization methods which were fairly similar in the titration curve are interval normalization and normalization by the total intensity. Normalization by interval has two major drawbacks, the first of which lies in the assumption that this interval is in fact correct. This is not necessarily a given, especially in the double mutants where the influence of the PRI effect needs to be considered. Additionally this approach is hard to standardize since it relies

on visual inspection and decision which areas to consider in the determination of the correction factor. Outliers in the chosen region, wrong choice of region and the assumptions about the interval are detrimental to the expected accuracy of the method.

Normalization by the total intensity avoids those issues by limiting the number of assumptions and manual steps, however this approach is only valid under the assumption that the relationship between the heights of individual peaks and the collective intensity is fixed. This turns out to be problematic either in case the assigned peaks are not identical between measurements or in case some phenomenon alters the ratio between the height of individual peaks and the overall intensity, like in this case the loss of PRI with increasing urea concentration might. Additionally it is only useful to even out errors across various samples or measurements of the same mutant but not across different mutants. Normalization by a region of constant intensity aims to remove the need for the PRI to be negligible in the normalization by only using regions unaffected by the PRI, however this re-introduces issues we faced with interval normalization and is therefore not ideal either.

The values obtained without interval normalization were the most pronounced, i.e. both m-value as well as ΔG_0 were the lowest. Values obtained without normalization fall in the middle range when compared to those obtained with various normalization procedures. Based on the titration curves normalization by total intensity as well as interval appear to be the most promising, however based on the m-value analysis no such trend is apparent. We believe that the approximative approach does hold some value, however the normalization or need therefore will have to be analysed further to make it more reliable.

6.5 PRI based on T_2 measurements

The PRI effect determined based on T_2-relaxation measurements provides much more reliable information than the approximation. In this study we used the PRI measurements to further the understanding of the solution structure of OPN. So far it had been known that OPN samples compacted states and that there is a conserved compacted region in the middle part of the protein. This region shows significantly increased PRI in all mutants which exhibit negative PRI. Based on

the PRI effect observed we suspect that the C-terminal part, as represented by OPN220 mutants containing C247, is the least compacted part of the protein. The fact that in the mutant C54-C247 significant positive PRIs are observed throughout the entire protein leads us to believe that there is no compacted state that includes both the N-terminus as well as the C-terminus of OPN220. This hypothesis is furthermore supported by the positive PRI found in mutant C188-C247, however the lack of any significant PRI in OPN220 C108-C247 outside the compacted core region, ranging from approximately positions 120-200, combined with the significant negative PRI in this region suggests involvement of the C-terminal region in the compacted core towards its N-terminal side.

The most pronounced (negative) PRI can be observed in OPN220 C108-C188. This is of little surprise considering that both those positions are inside or at least in close proximity to the suggested compacted core of the protein. Both mutants containing a mutation in position 54 as well as one in a central position (C54-C108 and C54-C188) show significant negative PRI. If we were to explain the observed effect in analogy to the hypothetical IDP described in fig. 13 this data would suggest the presence of states I and II but absence of state III. This appears to be a novel and important insight into the dynamic solution structure of OPN, which might be a very relevant piece of information in functional studies.

7 Conclusion

The study of IDPs has become a hot topic in structural biology in recent years but remains challenging for experimental characterization. NMR is one of the leading techniques in the analysis of IDPs with PRE measurements being a popular choice.

By exploiting a novel dipole-dipole electron-nucleus cross-correlation term, $\Delta^1\mathrm{H}^\mathrm{N} - \Gamma_2$ or PRI, we have managed to use this popular approach in a new way which provides a way to characterize transient compacted states of IDPs in more detail and gain insight into the global structural arrangement. We have applied the method to OPN and found that interactions of the N- and C-terminus with the compacted central region occur intermittently rather than together, therefore providing a first insight into the structure of OPN from the standpoint of a structural ensemble. The sigmoidality of the denaturation profile we observed clearly indicates that in spite of its flexibility OPN can by no means be described by a random coils and even undergo cooperative folding. The fact that the calculated $\Delta\mathrm{G}_0$ values are noticeably small supports the idea that rather than unfolded it is merely less rigidly stabilized than natively folded proteins and therefore rapidly samples structural variants. Overall this suggests that IDPs have much more in common with natively folded proteins than was originally thought and we should neither neglect their compaction nor dynamics in our quest to understand them.

While we have chosen to apply our technique to an IDP and found it to be promising in that challenging area of research the method can be used much more versatilely. It is by no means limited to the study of this set of proteins, but rather shows the potential to be of great interest in the study of all proteins and with some adaptation potentially even broader applications.

References

[1] P. E. Wright and H. J. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *Journal of Molecular Biology*, vol. 293, pp. 321–331, Oct. 1999.

[2] V. Ozenne, R. Schneider, M. Yao, J.-r. Huang, L. Salmon, M. Zweckstetter, M. R. Jensen, and M. Blackledge, "Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution," *Journal of the American Chemical Society*, vol. 134, no. 36, pp. 15138–15148, 2012.

[3] H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions," *Nature Reviews Molecular Cell Biology*, vol. 6, pp. 197–208, Mar. 2005.

[4] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, "Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling," *Journal of molecular recognition: JMR*, vol. 18, pp. 343–384, Oct. 2005.

[5] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker, "Sequence complexity of disordered protein," *Proteins: Structure, Function, and Bioinformatics*, vol. 42, no. 1, pp. 38–48, 2001.

[6] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, and A. K. Dunker, "Intrinsic disorder and functional proteomics," *Biophysical journal*, vol. 92, no. 5, pp. 1439–1456, 2007.

[7] D. Eliezer, "Biophysical characterization of intrinsically disordered proteins," *Current opinion in structural biology*, vol. 19, no. 1, pp. 23–30, 2009. IDP, Experimental Method, Study IDPs, Techniques, Review.

[8] H. J. Dyson and P. E. Wright, "Unfolded proteins and protein folding studied by nmr," *Chemical reviews*, vol. 104, no. 8, pp. 3607–3622, 2004.

[9] D. Eliezer, "Characterizing residual structure in disordered protein states using nuclear magnetic resonance," in *Protein Folding Protocols*, pp. 49–67, Springer, 2006.

[10] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature Structural & Molecular Biology*, vol. 9, no. 9, pp. 646–652, 2002.

[11] S. A. Adcock and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins," *Chemical reviews*, vol. 106, no. 5, pp. 1589–1615, 2006.

[12] T. H. Click, D. Ganguly, and J. Chen, "Intrinsically disordered proteins in a physics-based world," *International Journal of Molecular Sciences*, vol. 11, pp. 5292–5309, Dec. 2010.

[13] L. W. Fisher, A. Jain, M. Tayback, and N. S. Fedarko, "Small integrin binding ligand n-linked glycoprotein gene family expression in different cancers," *Clinical Cancer Research*, vol. 10, no. 24, pp. 8501–8511, 2004.

[14] A. Bellahcène, V. Castronovo, K. U. Ogbureke, L. W. Fisher, and N. S. Fedarko, "Small integrin-binding ligand n-linked glycoproteins (siblings): multifunctional proteins in cancer," *Nature Reviews Cancer*, vol. 8, no. 3, pp. 212–226, 2008.

[15] L. Fisher, D. Torchia, B. Fohr, M. Young, and N. Fedarko, "Flexible structures of sibling proteins, bone sialoprotein, and osteopontin," *Biochemical and biophysical research communications*, vol. 280, no. 2, pp. 460–465, 2001.

[16] P. Y. Wai and P. C. Kuo, "The role of osteopontin in tumor metastasis," *Journal of Surgical Research*, vol. 121, no. 2, pp. 228–241, 2004.

[17] D. R. Senger, D. F. Wirth, and R. O. Hynes, "Transformed mammalian cells secrete specific proteins and phosphoproteins," *Cell*, vol. 16, no. 4, pp. 885–893, 1979.

[18] H. Rangaswami, A. Bulbule, and G. C. Kundu, "Osteopontin: role in cell signaling and cancer progression," *Trends in cell biology*, vol. 16, no. 2, pp. 79–87, 2006.

[19] M. K. El-Tanani, F. C. Campbell, V. Kurisetty, D. Jin, M. McCann, and P. S. Rudland, "The regulation and role of osteopontin in malignant transformation and cancer," *Cytokine & growth factor reviews*, vol. 17, no. 6, pp. 463–474, 2006.

[20] V. Lamour, M.-J. Nokin, A. Henry, V. Castronovo, and A. Bellahcène, "Sibling proteins: molecular tools for tumor progression and angiogenesis," *Medecine sciences: M/S*, vol. 29, no. 11, pp. 1018–1025, 2013.

[21] K. X. Wang and D. T. Denhardt, "Osteopontin: role in immune regulation and stress responses," *Cytokine & growth factor reviews*, vol. 19, no. 5, pp. 333–345, 2008.

[22] S. A. Lund, C. M. Giachelli, and M. Scatena, "The role of osteopontin in inflammatory processes," *Journal of cell communication and signaling*, vol. 3, no. 3-4, pp. 311–322, 2009.

[23] C. M. Giachelli and S. Steitz, "Osteopontin: a versatile regulator of inflammation and biomineralization," *Matrix Biology*, vol. 19, no. 7, pp. 615–622, 2000.

[24] S. Ashkar, G. F. Weber, V. Panoutsakopoulou, M. E. Sanchirico, M. Jansson, S. Zawaideh, S. R. Rittling, D. T. Denhardt, M. J. Glimcher, and H. Cantor, "Eta-1 (osteopontin): an early component of type-1 (cell-mediated) immunity," *Science*, vol. 287, no. 5454, pp. 860–864, 2000.

[25] A. C. Renkl, J. Wussler, T. Ahrens, K. Thoma, S. Kon, T. Uede, S. F. Martin, J. C. Simon, and J. M. Weiss, "Osteopontin functionally activates dendritic cells and induces their differentiation toward a th1-polarizing phenotype," *Blood*, vol. 106, no. 3, pp. 946–955, 2005.

[26] M. Ishijima, S. R. Rittling, T. Yamashita, K. Tsuji, H. Kurosawa, A. Nifuji, D. T. Denhardt, and M. Noda, "Enhancement of osteoclastic bone resorption and suppression of osteoblastic bone formation in response to reduced mechanical stress do not occur in the absence of osteopontin," *The Journal of experimental medicine*, vol. 193, no. 3, pp. 399–404, 2001.

[27] D. Kurzbach, G. Platzer, T. C. Schwarz, M. A. Henen, R. Konrat, and D. Hinderberger, "Cooperative unfolding of compact conformations of the intrinsically disordered protein osteopontin," *Biochemistry*, vol. 52, no. 31, pp. 5167–5175, 2013.

[28] G. Platzer, A. Schedlbauer, A. Chemelli, P. Ozdowy, N. Coudevylle, R. Auer, G. Kontaxis, M. Hartl, A. J. Miles, B. A. Wallace, *et al.*, "The metastasis-associated extracellular matrix protein osteopontin forms transient structure in ligand interaction sites," *Biochemistry*, vol. 50, no. 27, pp. 6113–6124, 2011.

[29] G. M. Clore and J. Iwahara, "Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes," *Chemical reviews*, vol. 109, no. 9, pp. 4108–4139, 2009.

[30] A. Carrington and A. D. McLachlan, *Introduction to magnetic resonance: with applications to chemistry and chemical physics*. Harper & Row New York, 1969.

[31] J. Iwahara, C. Tang, and G. Marius Clore, "Practical aspects of 1h transverse paramagnetic relaxation enhancement measurements on macromolecules," *Journal of Magnetic Resonance*, vol. 184, pp. 185–195, Feb. 2007.

[32] I. Solomon, "Relaxation processes in a system of two spins," *Physical Review*, vol. 99, no. 2, p. 559, 1955.

[33] J. Iwahara, C. D. Schwieters, and G. M. Clore, "Ensemble approach for nmr structure refinement against 1h paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule," *Journal of the American Chemical Society*, vol. 126, no. 18, pp. 5879–5896, 2004.

[34] G. Otting, "Protein nmr using paramagnetic ions," *Annual review of biophysics*, vol. 39, pp. 387–405, 2010.

[35] A. Schedlbauer, R. Auer, K. Ledolter, M. Tollinger, K. Kloiber, R. Lichtenecker, S. Ruedisser, U. Hommel, W. Schmid, R. Konrat, *et al.*, "Direct

methods and residue type specific isotope labeling in nmr structure determination and model-driven sequential assignment," *Journal of biomolecular NMR*, vol. 42, no. 2, pp. 111–127, 2008.

[36] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, "Nmrpipe: a multidimensional spectral processing system based on unix pipes," *Journal of biomolecular NMR*, vol. 6, no. 3, pp. 277–293, 1995.

[37] T. Goddard and D. Kneller, "Sparkynmr assignment and integration software," *University of California*, 2006.

[38] J. L. Battiste and G. Wagner, "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data ," *Biochemistry*, vol. 39, pp. 5355–5365, May 2000.

[39] D. Kurzbach, A. Vanas, A. G. Flamm, N. Tarnoczi, G. Kontaxis, N. Maltar-Strmečki, K. Widder, D. Hinderberger, and R. Konrat, "Detection of correlated conformational fluctuations in intrinsically disordered proteins through paramagnetic relaxation interference," *Physical Chemistry Chemical Physics*, 2016.

8 Abbreviations

DEER	Double Electron-Electron Resonance
DMSO	Dimethylsulfoxide
DTNB	5,5'-Dithiobis(2-nitrobenzoic acid)
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethyldiaminotetraacidic acid
EPR	Electron Paramagnetic Resonance
IDP	Intrinsically Disordered Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MTSL	(1- oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethio-sulfonate
HMQC	Heteronuclear Multiple Quantum Coherence
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
OPN	Osteopontin
PAA	Polyacrylamide
PCR	Polymerase Chain Reaction
PCS	Pseudo Contact Shifts
PRE	Paramagnetic Relaxation Enhancement
PRI	PRE Interference
RDC	Residual Dipolar Coupling
RMSE	Root Mean Square Error
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOFAST-HMQC	Band-Selective Optimized Flip-Angle Short-Transient HMQC
TEMED	Tetramethylethylenediamine

9 Zusammenfassung

Hinweise auf Proteine, die in ihrer nativen Umgebung keine stabile tertiäre Struktur aufweisen, finden sich bereits seit den 1970er Jahren, doch erst etwa seit der Zeit der Jahrtausendwende gewinnt ihre Erforschung an Popularität. Seitdem hat sich gezeigt, dass derartige Proteine keineswegs Ausnahmeerscheinungen sind, sondern vielmehr einen signifikanten Anteil der bekannten Proteine ausmachen. Obwohl die Funktionalität dieser Proteine, die heute als intrinsisch ungefaltene Proteine bezeichnet werden, das in der Biologie lange Zeit als unumstößlich geltenden Struktur-Funktion-Paradigma in Frage stellen, steht diese heute außer Zweifel.

Ein solches intrinsisch ungefaltenes Protein ist Osteopontin (OPN), ein Mitglied der SIBLING (Small Integrin Binding LIgand, N-linked Glycoproteine) Proteinfamilie, das zentrale Funktionen in Biomineralisierung, Immunsystem und Signaltransduktion übernimmt. Das meiste Aufsehen hat Osteopontin jedoch durch seine Rolle in einer Vielzahl an Krebserkrankungen erlangt.

In dieser Arbeit stellen wir eine neue Kernspinresonanz (NMR) Methodik vor, die basierend auf paramagnetischer Relaxation Einblicke in die dynamische Struktur von intrinsisch ungefaltenen Proteinen erlaubt. Zu diesem Zweck werden OPN Varianten hergestellt, die ein oder zwei Cysteine enthalten. An die Seitenketten dieser Cysteine werden Nitroxid-Verbindungen angebracht, welche paramagnetisch sind. Da die Relaxation von Kernspins in der räumlichen Nähe dieser paramagnetischen Markierungen erhöht ist können daraus Informationen über die Faltung des studierten Proteins abgeleitet werden. Während üblicherweise der Effekt einer einzelnen paramagnetischen Probe untersucht wird basiert unsere neue Technik auf der Analyse der Relaxation in doppelt paramagnetisch markierten Proben, hinsichtlich kreuzkorrelierter Kernspin-Elektronenspin dipolarer Relaxation, die sich aus einer Interferenz der beiden paramagnetischen Zentren ergibt. Die Bestimmung dieses Effekts erfordert die Messung der Relaxation in einer doppelt markierten Probe, sowie zwei Proben, die jeweils eine dieser Markierungen enthalten. Diese Werte können entweder direkt durch Messung der transversalen Relaxation oder indirekt über eine Approximation basierend auf den beobachteten Intensitäten in zweidimensionalen Experimenten bestimmt werden. Die Detektion des Effekts erfolgt über die Differenz der paramagnetischen Relaxation bedingt

durch die beiden paramagnetischen Markierungen an sich - beobachtet in den beiden einfach markierten Proben - und jene, die bei Kombination der beiden Markierungen - in der doppelt markierten Probe - auftritt. Da dieser Effekt stark abstandsabhängig ist ermöglicht er die selektive Untersuchung ausschließslich kompaktierter Zustände, durch seine Winkelabhängigkeit erlaubt er die Beobachtung konzertierter Bewegungen im Protein. Die starken Mittelungseffekte, die bei der Messung intrinsisch ungefaltener Proteine durch deren strukturelle Vielfalt und Dynamik auftreten, erlauben keine

Wir illustrieren diese neu entwickelte Methodik anhand von OPN und gewinnen im Zuge dessen zugleich neue Informationen über die Struktur und Dynamik des Proteins. Anhand des Interferenzeffekts ist es uns gelungen zu zeigen, dass OPN in kompakten gefaltenen Strukturen existiert, wobei diese Strukturen eine zentrale Region des Proteins, sowie den N-Terminus oder C-Terminus, jedoch nicht beide Termini zugleich umfasst. Durch eine Harnstofftitration haben wir bestimmt, dass signifikante kooperative Faltung auftritt, die jedoch nur mit geringer Stabilität einhergeht. Diese Beobachtungen können als Indiz dafür verstanden werden, dass zumindest OPN nativ gefaltenen Proteinen strukturell im Grunde sehr ähnlich ist. Der entscheidende Unterschied liegt unseren Daten zu Folge in einer weniger rigiden Stabilisierung der Strukturen, die durch die geringere Aktivierungsenergie bei Strukturänderung erhöhte Dynamik des Proteins bedingt.