



DISSERTATION

Titel der Dissertation

Structural, Dynamic and Thermodynamic
Description of IDPs and their Interactions by NMR

Verfasser

Thomas Schwarz, BSc MSc

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2015

Studienkennzahl lt.
Studienblatt:

A 094 490

Dissertationsgebiet lt.
Studienblatt:

Molekulare Biologie

Betreut von:

Prof. Dr. Robert Konrat

Papers Used

This cumulative dissertation is based on the following publications:

Sara, T., Schwarz, T.C., Kurzbach, D., Wunderlich, C.H., Kreutz, C., Konrat, R. Magnetic resonance access to transiently formed protein complexes. *Chemistry Open* **3**, 115-123 (2014)

Kurzbach, D.,* Schwarz, T.C.,* Platzer, G., Hoefler, S., Hinderberger, D., Konrat, R. Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex. *Angewandte Chemie International Edition* **53**, 3840-3843 (2014)

*both authors contributed equally

Platzer, G.,* Kurzbach, D.,* Schwarz, T.C., Henen, M.A., Konrat, R., Hinderberger, D. Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin. *Biochemistry* **52**, 5167-5175 (2013)

*both authors contributed equally

Author Contributions

All authors contributed to the ‘Introduction’ and ‘Discussion’ parts of all manuscripts while the ‘Materials and Methods’ sections as well as the graphs used were produced by the person providing the respective experimental data. The author of this thesis, Thomas C. Schwarz, provided the following specific additional contributions:

Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin. *Biochemistry* **2013**: Production of protein as well as measurements and data treatment for the urea denaturing gradient, differential PREs (NaCl and urea) and chemical shift difference due to NaCl.

Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex. *Angewandte Chemie International Edition* **2014**: Protein production as well as measurements and data treatment for differential PREs of all four double mutants (the measurements of the free form were repeats of earlier measurements carried out to ensure consistency between measurements with and without heparin). Dynamic measurements (R_1 , R_2 and hetNOE) at two field strengths for the bound and unbound forms of Osteopontin. Protein production and measurements for the ITC data (together with a bachelor student). Treatment of ITC-data and preparation of graphs. Chemical shift, PRE, ITC and DOSY measurements in high salt conditions.

Magnetic Resonance Access to Transiently Formed Protein Complexes. *Chemistry Open* **2014**: Protein production, as well as measurement and data treatment and generation of graphs for all differential PRE and differential spin relaxation (R_1 , R_2 and hetNOE) data. All mentioned ITC-data. In addition to all figures directly related to the data mentioned above, the figure for the table of content was produced by the author.

Abstract

As a stable three dimensional structure of compounds that can catalyze biochemical reactions was already inferred in 1894 (Fischer), their chemical nature was elucidated in 1926 (Summer) and a stable fold was demonstrated as early as 1958 (Perutz, Kendrew), the requirement of stable structures to convey functions became a central dogma in molecular biology. Due to missing information on the abundance of highly flexible proteins and the fact that they are rarely involved in catalysis, they have been largely disregarded. Today, however, we know about their involvement in numerous physiological and pathological processes which has led to a dramatic increase in scientific interest.

Here we demonstrate state of the art magnetic resonance methods to characterize the dynamic, thermodynamic and structural properties of intrinsically disordered proteins (IDPs) and their interactions with the example of Osteopontin (OPN). This highly charged protein fulfills various different functions *in vivo* and is overexpressed in several types of tumor cells, where its interaction with receptors at the cell-surface (i.e. Integrins, CD44) are implicated in its tumorigenicity. This makes OPN not only a valuable model, but also a research target of high importance.

Our data show that OPN, despite being highly flexible and mostly extended, also samples well defined cooperatively folded compact conformations that show strong resistance to denaturing agents. Using new magnetic resonance methodology in conjunction with isothermal titration calorimetry (ITC), we are also able to provide thermodynamic information on the interaction of OPN with its naturally occurring ligand heparin. Here we detect increased amounts of disorder and relate them to the observed mechanism of interaction.

Zusammenfassung

Da auf eine stabile dreidimensionale Struktur von Verbindungen, welche biochemische Reaktionen katalysieren können, schon im Jahr 1984 (Fischer) rückgeschlossen wurde, ihre chemische Beschaffenheit 1926 (Summer) aufgeklärt wurde und die Faltung selbst bereits 1958 (Perutz, Kendrew) gezeigt wurde, ist das Vorhandensein einer stabilen Struktur zur Vermittlung einer Funktion ein zentrales Dogma der Molekularbiologie geworden. Fehlende Informationen zur relativen Häufigkeit extrem flexibler Proteine und deren seltene Involvierung in katalytischen Prozessen führten zu Ihrer Vernachlässigung in der Forschung. Heute jedoch ist ihrer Beteiligung an einer Vielzahl physiologischer und pathologischer Prozesse bekannt, was zu einem drastischen Anstieg des wissenschaftlichen Interesses geführt hat.

In dieser Arbeit demonstrieren wir modernste Methoden der Magnetresonanz zur Charakterisierung der dynamischen, thermodynamischen und strukturellen Eigenschaften intrinsisch ungeordneter Proteine (IDPs) und ihrer Interaktionen anhand des Beispielsproteins Osteopontin (OPN). Dieses stark geladene Protein erfüllt viele verschiedene Funktionen *in vivo* und wird in unterschiedlichsten Arten von Tumorzellen überexprimiert. Hier tragen seine Interaktionen mit Rezeptoren der Zelloberfläche (z.B. Integrine, CD44) zur Kanzerogenität bei. Daher ist OPN nicht nur ein wertvolles Modellsystem, sondern auch ein hoch relevanter Forschungsgegenstand.

Unsere Daten zeigen, dass OPN trotz seiner großteils extendierten Form und seiner hohen Flexibilität auch gut definierte, kooperativ gefaltete, kompakte Zustände populiert und diese auch gegenüber denaturierenden Stoffen Widerstand zeigen. Durch die Nutzung neuer Magnetresonanzmethodologie in Verbindung mit isothermaler Titrationskalorimetrie können wir auch thermodynamische Information zur Interaktion zwischen OPN und seinem natürlich vorkommenden Liganden Heparin sammeln. Hierbei messen wir erhöhte Mengen an Unordnung und stellen diese mit dem beobachteten Interaktionsmechanismus in Verbindung.

Table of Contents

<u>Introduction</u>	<u>11</u>
History	13
Structure	14
NMR for IDPs	16
Protein Interactions	19
Protein Dynamics	22
Osteopontin	24
Overview	25
References	26
<u>Chapter 1</u>	<u>37</u>
Introduction	39
Experimental Procedures	41
Results and Discussion	44
Conclusion	58
References	60
Supporting Information	65
<u>Chapter 2</u>	<u>73</u>
Results and Discussion	75
References	83
Supporting Information	85
<u>Chapter 3</u>	<u>97</u>
Introduction	99
Results and Discussion	101
Conclusions	118
References	120
<u>Discussion</u>	<u>125</u>
Chapter 1	128
Chapter 2	130
Chapter 3	134
Related Work	137
Concluding Remarks	141
References	144
<u>Acknowledgements</u>	<u>149</u>
<u>Curriculum Vitae</u>	<u>153</u>

Introduction

History

All organisms rely on an array of different compounds in order to generate all the chemical reactions we commonly call 'life'. These compounds are classified in a variety of ways and include, among many others, lipids, proteins, fatty acids, sugars, DNA and RNA. Although enzymes were known to play important roles in many biological processes (i.e. fermentation), it took researchers until 1926 to verify that proteins are the actual constituents carrying out enzymatic reactions.¹ Interestingly, even then, the protein sample was obtained by crystallization, as it was the method of choice to prepare a clean sample of any unknown molecule.

Up until 1952, when Hershey & Chase² were able to verify previous speculations³ that DNA was the constituent of biological systems, that enables any organism to pass on its genetic information, proteins were often assumed to be the hereditary material. This speculation was due to the fact, that proteins consist of 20 different amino acids, and DNA consists only of 4 different bases, more information could theoretically be stored in less material. The idea that this higher theoretical information content *must* mean that proteins carry the genetic information was so widespread, that it persisted up until 1952, even though proteins had already been shown to be the constituent catalyzing at least some chemical reactions.¹

Due to very insightful chemistry they were already presumed to need a well-defined structure in order to carry out these reactions.⁴ First direct insights into these structures themselves came with the advent of X-Ray crystallography.^{5,6,7} These first examples of how a long and seemingly random chain of amino acids can adopt such a well defined three dimensional structure were a great leap forward in understanding protein function on a molecular level. As more and more structures were solved with this technology, the belief that all proteins necessarily need to have a single well defined structure in order to carry out their function became deeply entrenched in most researchers minds and if proteins did not display a stable structure, this aspect was often simply ignored.⁸ This static picture of protein structure even persisted after first studies demonstrated, that some proteins that have clear and important functions do not display a stable fold in solution.⁹ It was merely altered to include proteins that do not have a stable structure when free in solution, but still adopt a well defined fold in their active state (i.e. 'folding upon binding').

Introduction

As the ‘active’ state of the protein was still perceived as structured, investigations into the conformational ensemble the proteins displays before interaction were not regarded as highly important. In addition to the theoretical unimportance of proteins that do not display a single stable structure, there were methodological obstacles preventing the investigation of proteins that do not fold into a single structure. X-ray diffraction depends on the constructive interference of individual diffractions, which is caused by atoms only if they are arranged in a protein crystal.¹⁰ If there are many conformations present, crystal formation is virtually impossible and there is also no possibility for a clear diffraction pattern, thus preventing any investigations into the structure of the protein. Elimination of flexible parts within proteins was often used as a tool to help crystallization, neglecting the importance of unstructured regions.^{11,12} As the wealth of X-ray crystallographic data grew, the idea of folded proteins being the only relevant state for biological function became ever more of a dogma. Although the very first hints towards unstructured proteins having very relevant functions were discovered much earlier,⁹ the broader interest in these proteins only came when bioinformatic analyses were able to demonstrate that a large part of our proteome is at least partially unstructured in its ‘natural’ state.¹³

Structure

As more and more structures of proteins were solved and many full genomes were sequenced, bioinformatic disorder predictors could be developed and refined to high accuracy. In the case of DISOPRED2¹⁴ a false positive rate of only 0.5% was achieved on the test set. Even this number is probably overestimated, as some of the predicted disordered regions showing order in the crystal structure do so due to contacts with binding partners or the crystal lattice. Using this program, it was estimated that stretches of thirty or more consecutive disordered amino acids occur in 2% of the archaean, 4.2% of eubacterian and 33% of the eukaryotic genome.¹⁴ As there is a significant likelihood of disordered regions being falsely predict as ordered (due to the training of the algorithm on crystal structures) this is most likely a conservative estimation.

Additionally, the advent of fast and easy sequencing technologies has led to a dramatic shift in how proteins are investigated. Previously purified fractions of proteins were tested for (mostly enzymatic) activity and subsequently purified and only then was their amino acid composition identified. These methods mostly detected well-folded proteins. As sequencing became ubiquitous many researchers began to start with a given sequence whose product has been shown to be important for a specific cell function and subsequently focus on the protein it encodes. As this methodology of the so-called ‘post-genomic era’ does not place as many requirements on the structure of the protein, it led to the discovery of IDP abundance.¹⁵

Once it became apparent that a large part of the proteome is intrinsically disordered, while theoretical concepts still explained protein functions in terms of a single stable structure, the question commonly asked was ‘Why are so many proteins not adopting a stable tertiary fold?’. Investigations into the advantages of IDPs lead to the discovery of various attributes that are beneficial in the right context. The flexible nature of IDPs can allow for remarkably fast on-rates, possibly facilitated by a ‘fly-cast’ like mechanism.^{16,17} Sampling of many different conformations can allow for specific interactions with different binding partners while folding upon binding, which is observed for many IDPs, can potentially lower the binding affinity. Relatively low affinities are sometimes a requirement for interactions of high specificity to properly function in signal transduction networks.¹⁸ One part of the protein might be involved in an interaction, whilst another part can still sample a relatively large area and is not perturbed in its interaction with a second binding partner, thus enabling IDPs to function as ‘scaffold-proteins’.¹⁹ Additional explanations on the usefulness of IDPs can be found when heading Dobzhansky’s famous quote: “Nothing in biology makes sense except in the light of evolution.”²⁰ Not only does the use of IDPs greatly expand the accessible sequence space for proteins, the relative stability of unstructured proteins towards mutations as well as their adoption of many different conformations means that new functions can be evolved more easily for these proteins in comparison to their folded counterparts.^{21,22} In addition the question mentioned above (‘Why are so many proteins not adopting a stable tertiary fold?’) is probably not the most helpful approach to IDPs, as it still stems from the view of a dichotomic partitioning of proteins into ‘ordered’ versus ‘disordered’. Research regarding the excited states of proteins has demonstrated that they can also contain structures within their ensemble

which differ greatly from the observed ground state structure.²³ Keeping this in mind, the difference between IDPs and ‘folded’ proteins is reduced to variations in the number of thermally accessible states stemming from differing energy landscapes. As more and more IDPs are shown to contain populations of (at least partly) folded conformers, while the excited states of ‘folded’ proteins reveal their ability of adopting varying structures,²⁴ the clear distinction between ‘folded’ versus ‘unfolded’ fades away. Rather than being the only two relevant versions of structure, they should be viewed as two extremes in a continuum.²⁵

NMR for IDPs

Fortunately the limitation to well folded proteins (as in X-ray crystallography) does not apply for all spectroscopic methods. The rapid interconversion between numerous different states, however, poses a challenge for all structure determination methods. Solution based methods commonly can only access a time averaged structural ensemble of an IDP. Using small angle X-ray scattering (SAXS) the general shape of the ensemble as well as the distance distribution expected can be calculated and recent advances can even provide insights in the overall shape of IDP ensembles.²⁶ The introduction of two fluorescent labels allows for measurement of the energy transfer efficiency between them, thus enabling the calculation of distance distribution in Förster resonance energy transfer (FRET) experiments.²⁷ In both of these methodologies different distances contribute with different signal intensity, the rapid interconversion between several states with ever changing distances, thus, complicates the determination of accurate distance distributions. In the case of circular dichroism (CD) measurements the determination of the secondary structure content is generally achieved by an approximation of the observed curve with theoretical curves of pure secondary structures, often using neural networks.²⁸ The estimation of average secondary structure content is, thus, possible but information on the contributions of individual conformations to the average is lacking. Therefore a distinction of whether 10% of the protein is α -helical 90% of the time or 90% of the protein 10% of the time is not possible. Using unpaired electrons as spin labels, continuous wave (CW) electron paramagnetic resonance (EPR) can report on the mobility of the introduced spin label in the context of the full protein.^{29,30}

Technological advances such as the development of ultra high field instruments and multidimensional experiments,^{31,32} together with improved protein production and labeling schemes^{33,34} have helped nuclear magnetic resonance (NMR) spectroscopy make significant progress towards a description of ever larger proteins without a single stable fold. Although many other methods can be used to gather information on the conformational ensemble of these proteins, NMR stands out due to both the wealth of different types of information it can provide and its capability of accessing this information on an atomic level. Obviously this methodology also comes with its limitations: Although ever bigger structures are solved by NMR, the size limitation is still the major obstacle preventing NMR from becoming the method of choice for most structure determination efforts. In the case of IDPs, however, the size limitation stemming from the slow tumbling times of the protein is less stringent as high motional freedom of the backbone leads to slower relaxation. The effect of peak overlap is worse in proteins without a defined tertiary structure, but experiments of high dimensionality and direct carbon detection can be used to circumvent this problem.³⁵ The need for isotope labeled samples leads to a restriction in expression systems, which makes it difficult to access human proteins with the inclusion of their post translational modifications, although the introduction of these modifications to purified protein extracts is often possible.³⁶ The requirement for a sample which can be highly concentrated and is stable for extended periods of time close to room temperature is also often a hard obstacle to overcome. In most cases, however, this can be achieved after testing for permissive buffer conditions and additives. The relatively high vulnerability of unstructured segments in terms of protease activity *in vitro* often complicates investigations, but it can also be put to good use when using differential protease cleavage patterns of bound and unbound species to identify interaction sites.³⁷ In contrast to the *in vitro* observations, *in vivo* the degradation rate of IDPs in general does not seem to be drastically higher than observed for fully folded proteins.^{38,39}

A large set of different measurement techniques has been developed by NMR-spectroscopists to access the conformations of proteins in general and IDPs in particular. First and foremost signal assignment needs to be carried out in order to determine the origin of the resonances observed. As signal assignment in proteins is typically achieved along the proteins backbone, C α and C β chemical shifts are typically collected alongside carbonyl carbon, amide nitrogen and proton chemical shifts. As the deviation of these chemical shifts from statistical random coil values are

Introduction

indicative of secondary structures, a first step is often to calculate the secondary structure propensity of each residue of the protein under investigation.⁴⁰ This propensity, however, does only give information on local structure formation. When looking for the overall structure of proteins, information on long-range contacts between amino acids is of paramount importance. In folded proteins the nuclear Overhauser effect (NOE) is typically used to provide information on such contacts. In IDPs the small populations of each state together with the distance dependence of r^{-6} for the NOE call for a different way of investigation. The method of choice in this respect is the measurement of enhanced relaxation rates caused by the introduction of a paramagnetic spin label (PREs). As the paramagnetic label causes a decrease in signal intensity in a distance dependent manner, monitoring the changes along the proteins primary sequence allows for the determination of transient contacts and compactions within the protein.⁴¹ In the context of folded proteins PREs stemming from a paramagnetic environment are a very useful tool to gain additional structural information, this is especially true for large proteins that require deuteration and provide much less NOE information.⁴² Additional experiments for structural determination include the measurement of (cross-correlated) backbone dihedral angle scalar couplings⁴³ and the measurement of solvent exchange rates and residual dipolar couplings (RDCs).^{44,45} The last two techniques are very useful in well folded proteins but, due to the low signal dispersion in IDPs and the distribution of their torsion angles in the ramachandran plot, the utility of cross correlation experiments for IDPs has only recently come to light.⁴³ The relatively high amounts of alignment required to obtain RDC measurements of an intrinsically disordered protein, on the other hand, can easily lead to a perturbation of the conformational ensemble. This possibility needs to be excluded very carefully using multiple alignments and complementary methods.⁴⁶ Therefore, care has to be taken when using these techniques and they are generally not used as much when studying IDPs.

As IDPs are defined by their inability to adopt a stable tertiary fold, descriptions of their structure can never be performed with a simple set of atomic coordinates. Descriptions must, therefore, always contain a relatively high inaccuracy in regards to atomic positions making them 'low-resolution' structures. However, even what became to be the most famous three-dimensional structure of a biological macromolecule would be considered a low-resolution, integrative structure today. The structure of the DNA double helix as published by Watson and Crick⁴⁷ was based on

low-resolution fiber diffraction data as well as chemical constraints and stoichiometry considerations, nevertheless the conclusions drawn from this model led to a flurry of new techniques and investigations in what are now called the life-sciences. In addition the concept of ‘resolution’ implies a static model, but as structure and dynamics heavily influence each other, this is particularly problematic for all flexible macromolecules. Therefore, we can conclude that it is not the ‘resolution’ of a certain model, but the conclusions drawn from it and the new hypothesis that can be tested, which determine its usefulness.⁴⁸

Protein Interactions

Early investigations to discern how yeast was able to specifically digest one stereoisomeric form of sugar, led to the first model for protein interactions. The first structural implications of how proteins (or active chemical agents “active chemische Agentien”, as Fischer called them) are involved in these interactions were imagined in 1894, when Fischer proposed the lock-and-key model.⁴ Here the protein and the ligand were viewed as rigid objects that interact due to their shape and the properties of their surface. No change in their conformation (or ‘shape’) is expected during the interaction. This model did explain some key elements of protein-ligand interactions like specificity and recognition. The stabilization of intermediates within the active enzyme, which is often required in enzymatic processes, however, could not be explained with this model. A less static picture of interactions was achieved with the ‘induced-fit’ model, postulated in 1958 by Koshland.⁴⁹ The urge to develop this theory came from the observation of rapid synthesis of a complex polymer with defined chemical units at specific positions (i.e. a proteins). To explain the rapid reaction and lack of intermediates containing only some of the product, the model of a template together with a single enzyme was proposed (with the possibility of the template being the enzyme). For a single enzyme to react in a specific order with a set of different amino acids, the ‘induction’ of a new template area after each peptide bond formed, was postulated. This theory encompasses slight changes in both interacting partners, upon interaction. With this model the generation of new ‘active-sites’ and, thus, the rapid synthesis of proteins could be explained. Interestingly, all the tremendous amount of progress in the field, which did yield a much more detailed description of

Introduction

the ribosome, did not change basic idea of the active site model.⁵⁰ Additionally, it offered an answer as to why enzymes sometimes cannot process compounds, which are analogous to their substrate but slightly smaller. For example alpha methyl glucoside should act as a substrate for amylomaltase, but does not.⁵¹

Although the model of an ‘induced-fit’ did encompass structural adaptation of the interacting partners, they were still seen as having a single most stable state, which is modified upon interaction. As proteins were still viewed as relatively static objects, the question to be answered for newly found interactions was whether it was a ‘lock-and-key’ or an ‘induced-fit’ type of interaction. Very early on evidence started to accumulate that proteins are much less static than previously envisioned^{52,53,54} and the explanation of their interactions also needed a model that would allow for greater flexibility.

The first approximations of the folding behavior of proteins claimed that the polypeptide chain would form consecutively more and more contacts that maximized stabilizing forces (i.e. Van der Waals interactions, hydrogen-bonds and salt-bridges) while entropic costs stemming from increased order of water molecules around exposed hydrophobic surfaces would be minimized.^{55,56,57} Plotting the energy of each state along the y-axis and conformation along x should, therefore, give a smooth graph with a single minimum. More realistically, however, it has a rugged shape and contains many local minima. In the case of well-folded proteins most commonly a single global minimum is observed. However, local minima are present and can be populated to smaller degrees. This is due to the fact that structural fluctuations at ambient temperature are frequently large enough to overcome energy barriers between the global and local minima and the populations follow a thermodynamic distribution. In IDPs many different minima of roughly the same energy and without high barriers between them are observed, thus allowing for rapid interconversion between many different states.⁵⁹ The idea of a ‘protein folding funnel’^{58,60} also had consequences regarding our understanding of protein interactions. In this context it is expanded to the ‘energy landscape theory’.^{61,58,62} Here, interactions signify the selection of a certain sub-state and a concomitant reduction of its free energy upon interaction, which makes them greatly dependent on the structural ensemble. In the framework of ‘conformational selection’, interactions are viewed as a combination of the conformational energy landscapes of two proteins and the energy landscape of their interaction.⁶³ Here, the binding conformation of both proteins is already present in

their structural ensemble prior to interaction. The interaction itself simply occurs when the ‘correct’ conformation of both partners is encountered. As the ‘correct’ conformation of one or both interacting partners does not necessarily have to be a highly populated conformation, this type of interaction is often also seen as a selection of one ‘excited state’ of the protein. Recent advances in NMR methodology have provided insight in high-energy structures and have, for example, allowed for the detection of sparsely populated conformations in the free form of MBP. These conformations only represent 5-10% of the protein ensemble and resemble the maltotriose bound form of the protein.⁶⁴

As measurements of dynamics by NMR or single molecule methods revealed that in some interactions conformational selection is followed by dynamic adaptations,^{65,66} the model of conformational selection was extended to account for these possibilities.⁶⁷ In the extended view, both ‘lock and key’-type mechanisms, where two rigid bodies interact, and the ‘induced-fit’-type, where the interaction leads to slight conformational changes that allow for tight binding, are seen as extremes of the same mechanism: A combination of their energy landscapes for interaction and the possibility of the interaction itself changing the resulting energy landscape. Here, a change in the energy landscape subsequently leads to an adaptation of the proteins structure. In this way the ‘lock and key’ type of interaction, the ‘induced-fit’ type and any combination between them can be explained with the ‘extended conformational selection’ model. For example it encompasses the selection of a lowly populated substate out of an ensemble to form an encounter-complex followed by a change in the energy landscape and the subsequent population of a free state not found in solution. The description of such mixed processes is progressing rapidly.⁶⁸ The ‘extended conformational selection’ model also ameliorates concerns arising from the observation that shifts from a more ‘conformational selection’ to an ‘induced-fit’ type of behavior have been observed with increasing ligand concentrations.^{69,70} This is not to say that a generally accepted theory for protein interactions has already been found, as various models are still being discussed and the discussion is often based on the ‘original conformational selection’ rather than the ‘extended’ theory.^{71,67,72,73,74,75}

Additionally, the ‘extended induced-fit’ still assumes binding to be achieved through a single energy minimum being formed. Protein interactions that are achieved by one protein accessing various binding conformations (‘polymorphic’) or have

flexible regions that scan for electrostatic interaction ('fuzzy') have been described,⁷⁶ indicating that further adaptations to the theory will have to be made. In addition dynamic interactions that simultaneously make use of transient long-range contacts in extended conformations while using compact conformations to optimize the mean electrostatic field of interaction have been observed.⁷⁷

Protein Dynamics

The first molecular dynamics simulations of proteins were performed as early as 1977,⁷⁸ outside of the computational community the motions within proteins were generally viewed as small fluctuations from the coordinates, which are not relevant to the overall structure. Dynamics within proteins were even compared to those in dense materials with only some concerted motions from one state to another when required for biological function.⁷⁹ Averaged structures have helped to explain many fundamental questions in structural biology, but they do not capture proteins (and in particular IDPs) in their full complexity. In reality all macromolecules, and proteins in particular, display motions that range from nano- to micrometers and from femtoseconds to hours. Coming from the static view of protein crystallography, regions of proteins that were not observable in the crystal structure were (and still are) often described as 'dynamic', without further clarification. As many kinds of motions within proteins are necessary for functions ranging from binding^{80,76,81} to catalytic turnover,^{82,83} their detailed description in terms of position, amplitude and frequency is often paramount for a complete understanding of protein function.

Nuclear magnetic resonance spectroscopy is on the forefront of these investigations, as it has the unique capability of accessing many different motional timescales at an atomistic level.⁸⁴ Access to these motions is possible as the three fundamental observables in NMR (chemical shift, intensity and linewidth) can be sensitive to motions on various timescales. In contrast to other spectroscopic techniques (IR, CD, UV-Vis, etc.) NMR can use different isotope labeling schemes to selectively access nearly all atoms within the protein. As the isotopes used are nearly indistinguishable from their naturally more abundant counterparts, other than in regard to their molecular mass and nuclear spin, their incorporation introduces only minor

changes in the protein, which are not expected to influence its natural behavior and dynamics.

Various pulse sequences and methods of analysis have been developed to take advantage of the dynamic sensibility of NMRs observables. The range of motions accessible under the right circumstances has been gradually expanded and reaches from picosecond motions like side-chain rotations to minutes or even days for slow processes like hydrogen exchange rates of protons involved in hydrogen bonds of very stable proteins. Unfortunately not all timescales are accessible with ease and some require stably folded proteins and very laborious methodology,^{85,86} leading to the inaccessibility for motions in IDPs occurring in the intermediate nanosecond to the low microsecond range. Recent advances in instrumentation and experimental techniques have made motions in the range of roughly 20 to 100 μ s accessible to NMR without the requirement of partial alignment.⁸⁷ Thus, complete accessibility of all relevant motional timescales even for IDPs does seem possible in the future.

One method used to determine dynamic properties with NMR is the measurement of Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion, which will be explained in more detail in *Chapter 3*. This method is very powerful in generating information on lowly populated ($>0.75\%$) states, provided the exchange rate between the studied minor and the major state is on the ms to μ s timescale.⁸⁸ As exchange rates between substates in IDPs are commonly faster, information on their dynamics is usually obtained by measurement of the relaxation rates R_1 and R_2 of the amide ^{15}N as well as heteronuclear Overhauser effect (nOe) efficiency between ^{15}N and ^1H . These three observables depend on the prevalence of motions on different timescales (quantified by spectral density functions). In order to quantitatively analyze protein motions by NMR, the Lipari-Szabo formalism^{89,90,91} was introduced. It relates observed values for R_1 , R_2 and hetNOE with an order parameter S^2 (which represents the freedom of motion of the NH amide cone) and the corresponding correlation time τ . Extension of the formalism led to the possibility to dissect the order parameter into a very fast and a slower component.⁹² Further improvements like the calculation of the tumbling time and the possibility to use a non-spherical diffusion tensor have been implemented in freely available programs,^{93,94,95} which are nowadays routinely used in NMR order parameter assessment. Recent examples include the determination of relative domain movements not observed in X-ray diffraction,⁹⁶ the estimation of entropic contributions to binding mechanisms,⁹⁷ and the explanation of observed

regulation by changes in dynamics.⁹⁸ Investigations into functional dynamics can even correlate binding affinities and binding modes with dynamic behavior and use this information to modify binding properties with mutations remote to the binding site.⁹⁹

Osteopontin

Osteopontin (OPN) is a secreted chemokine-like protein and a member of the small integrin binding ligand, N-linked glycoprotein (SIBLING) protein family. OPN expression was demonstrated in several different tissues and cell types.^{100,101} OPN and other members of this protein family play important roles in many stages of cancer progression, as identified by gene transfer experiments.^{102,103} OPN and most of the other family members are highly over-expressed in several cancer cell lines and their abundance directly correlates with the tumor grade.^{104,105} Structurally, OPN is usually characterised as an intrinsically disordered protein as classical secondary structure predictors as well as meta-structure analysis indicate a lack of stable secondary structures. The predicted lack of structure is easily understood when looking at the primary sequence of OPN, which carries a high charge density, is depleted in hydrophobic amino acids and does not contain cysteines. In addition OPN has been shown to carry several post translational modifications (PTMs) such as glycosylation and phosphorylation.¹⁰⁶ The primary identification of OPN was obtained through rat bone demineralization.¹⁰⁷ It is present at high levels in the extracellular matrix (ECM) of bone, cementum and dentin¹⁰⁸ where it controls calcium mineralisation and is involved in the control of osteoblast attachment. Our groups' attention to this protein stems from the fact that it is overexpressed in avian fibroblasts when these are cooperatively transformed by the *v-myc* and *v-mil* onocogenes, which stem from the acute carcinoma virus MH2. The tumorigenic function of OPN is linked to its interaction with Integrin and CD44 receptors, which influences cell-signalling involved in cell motility, adhesion and survival.^{109,110} The interaction with this intricate signalling network leads to the involvement of OPN in tumor invasion, metastasis and angiogenesis.^{111,112} In addition its interaction with matrix metalloproteinases which affects extracellular matrix remodelling is key during tumor invasion.¹¹³

As Osteopontin is a so called 'IDP' of high biological relevance, studying its structure and dynamics can yield insight into both, the general properties of proteins

that lack a well defined tertiary structure, and into particular features of OPN relevant for its specific biological actions. Additionally it has favourable NMR features, which was beneficial to our investigation.

Overview

As outlined above, intrinsically disordered proteins (IDPs) constitute a large and diverse set of proteins with varying functions that are vital for eukaryotic life.^{80,76,81,82,83} Despite the mounting evidence of the importance of IDPs, their structural and dynamic characterization still lags far behind the one achieved for more structured proteins. In order to improve our understanding of the various properties of this protein class, we performed an in-depth analysis of its member Osteopontin.

Our protein of interest, Osteopontin, was recently analyzed using state-of-the-art NMR and other biophysical techniques.¹¹⁴ This publication already clearly established the existence of compact structures within the ensemble of OPN, using a variety of different techniques. Building on this publication **Chapter 1** focuses on the dynamic parameters of this structural ensemble in general and its more compact states in particular. Nuclear and electron magnetic resonance techniques are used to not only show the existence of more compact substates within the conformational ensemble of OPN, but also to unambiguously demonstrate the cooperative unfolding of these substrates. A detailed description of the interactions of OPN with its natural ligand heparin is carried out in **Chapter 2**. The characterization of this interaction does not only reveal distinct changes in the ensemble of OPN upon interaction, but it also demonstrates a possible mechanism for disordered proteins to lessen the entropic penalty that usually needs to be paid when binding of an IDP leads to its folding and, thus, its rigidification. A more general view of the techniques available to date for the structural and dynamic characterization of IDPs, especially in regard to their interactions, is given in **Chapter 3**. Here various NMR based techniques for the characterization of transiently formed complexes are outlined (demonstrated). Their relevance in regard to lowly populated (excited) states as well as in regard to IDPs is discussed and the examples provided clearly demonstrate the utility of these techniques and exemplify the conditions best suited for each approach.

References

- 1 Summer, J. B. The Isolation and Crystallization of the Enzyme Urease. *J Biol Chem* **69**, 435-441 (1926).
- 2 Hershey, A. D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **36**, 39-56 (1952).
- 3 Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *J Exp Med* **79**, 137-158 (1944).
- 4 Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Deut Chem Ges*, 2985-2993 (1894).
- 5 Bragg, W. L. & Perutz, M. F. The structure of haemoglobin: Fourier projections on the 010 plane. *Proc. R. Soc. Lond. Ser. A*. **225**, 315-329 (1954).
- 6 Kendrew, J. C. *et al.* Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **185**, 422-427 (1960).
- 7 Perutz, M. F. Stereochemistry of cooperative effects in hemoglobin. *Nature* **327**, 726-739 (1970).
- 8 Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, 553-584, doi:10.1146/annurev-biochem-072711-164947 (2014).
- 9 Boesch, C., Bindi, A., Oppliger, M. & Wuthrich, K. ¹H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur J Biochem* **91**, 209-214 (1978).
- 10 Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J* **280**, 5705-5736, doi:10.1111/febs.12495 (2013).
- 11 McPherson, A. *Preparation and Analysis of Protein Crystals*. (Wiley, 1982).
- 12 Dale, G. E., Oefner, C. & D'Arcy, A. The protein as a variable in protein crystallization. *J Struct Biol* **142**, 88-97 (2003).
- 13 Wootton, J. C. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem* **18**, 269-285 (1994).
- 14 Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-645, doi:10.1016/j.jmb.2004.02.002 (2004).

- 15 Linding, R. *et al.* Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453-1459 (2003).
- 16 Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**, 8868-8873, doi:10.1073/pnas.160259697 (2000).
- 17 Shammass, S. L., Travis, A. J. & Clarke, J. Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX. *J Phys Chem B* **117**, 13346-13356, doi:10.1021/jp404267e (2013).
- 18 Dunker, A. K. & Uversky, V. N. Signal transduction via unstructured protein conduits. *Nat Chem Biol* **4**, 229-230, doi:10.1038/nchembio0408-229 (2008).
- 19 Babu, M. M., van der Lee, R., de Groot, N. S. & Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol* **21**, 432-440, doi:10.1016/j.sbi.2011.03.011 (2011).
- 20 Theodosius, D., Vol. 35 125-129 (1973).
- 21 Stein, A., Pache, R. A., Bernado, P., Pons, M. & Aloy, P. Dynamic interactions of proteins in complex networks: a more structured view. *FEBS J* **276**, 5390-5405, doi:10.1111/j.1742-4658.2009.07251.x (2009).
- 22 Gerhart, J. & Kirschner, M. The theory of facilitated variation. *Proc Natl Acad Sci U S A* **104 Suppl 1**, 8582-8589, doi:10.1073/pnas.0701035104 (2007).
- 23 Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol* **5**, 808-814, doi:10.1038/nchembio.238 (2009).
- 24 Mittag, T. *et al.* Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **18**, 494-506, doi:10.1016/j.str.2010.01.020 (2010).
- 25 Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197-208, doi:10.1038/nrm1589 (2005).
- 26 Receveur-Brechot, V. & Durand, D. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci* **13**, 55-75 (2012).
- 27 Haas, E. The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements. *Chemphyschem* **6**, 858-870, doi:10.1002/Cphc.200400617 (2005).
- 28 Bohm, G., Muhr, R. & Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng* **5**, 191-195 (1992).
- 29 Hubbell, W. L., Gross, A., Langen, R. & Lietzow, M. A. Recent advances in site-directed spin labeling of proteins. *Curr Opin Struct Biol* **8**, 649-656 (1998).

- 30 Bordignon, E. Site-Directed Spin Labeling of Membrane Proteins. *Epr Spectroscopy: Applications in Chemistry and Biology* **321**, 121-157, doi:Doi 10.1007/128_2011_243 (2012).
- 31 Bermel, W. *et al.* Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. *J Biomol NMR* **55**, 231-237, doi:10.1007/s10858-013-9704-3 (2013).
- 32 Solyom, Z. *et al.* BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR* **55**, 311-321, doi:10.1007/s10858-013-9715-0 (2013).
- 33 Lichtenecker, R. J., Weinhaupl, K., Schmid, W. & Konrat, R. alpha-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *J Biomol NMR* **57**, 327-331, doi:10.1007/s10858-013-9796-9 (2013).
- 34 Lichtenecker, R. J., Coudevylle, N., Konrat, R. & Schmid, W. Selective isotope labelling of leucine residues by using alpha-ketoacid precursor compounds. *Chembiochem* **14**, 818-821, doi:10.1002/cbic.201200737 (2013).
- 35 Bermel, W. *et al.* Speeding up sequence specific assignment of IDPs. *J Biomol NMR* **53**, 293-301, doi:10.1007/s10858-012-9639-0 (2012).
- 36 Bah, A. *et al.* Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*, doi:10.1038/nature13999 (2014).
- 37 Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* **93**, 11504-11509 (1996).
- 38 Gsponer, J., Futschik, M. E., Teichmann, S. A. & Babu, M. M. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* **322**, 1365-1368, doi:10.1126/science.1163581 (2008).
- 39 Tompa, P., Prilusky, J., Silman, I. & Sussman, J. L. Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins* **71**, 903-909, doi:10.1002/prot.21773 (2008).
- 40 Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* **15**, 2795-2804, doi:10.1110/ps.062465306 (2006).
- 41 Clore, G. M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108-4139, doi:10.1021/cr900033p (2009).

- 42 Madl, T., Bermel, W. & Zangger, K. Use of relaxation enhancements in a paramagnetic environment for the structure determination of proteins using NMR spectroscopy. *Angew Chem Int Ed Engl* **48**, 8259-8262, doi:10.1002/anie.200902561 (2009).
- 43 Stanek, J., Saxena, S., Geist, L., Konrat, R. & Kozminski, W. Probing local backbone geometries in intrinsically disordered proteins by cross-correlated NMR relaxation. *Angew Chem Int Ed Engl* **52**, 4604-4606, doi:10.1002/anie.201210005 (2013).
- 44 Tjandra, N., Szabo, A. & Bax, A. Protein Backbone Dynamics and ¹⁵N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects. *J. Am. Chem. Soc.* **118**, 6986-6991 (1996).
- 45 Tjandra, N. & Bax, A. Measurement of Dipolar Contributions to ¹JCH Splittings from Magnetic-Field Dependence of JModulation in Two-Dimensional NMR Spectra. *J. Magn. Res.* **124**, 512-515 (1997).
- 46 Schwalbe, M. *et al.* Predictive atomic resolution descriptions of intrinsically disordered hTau40 and alpha-synuclein in solution from NMR and small angle scattering. *Structure* **22**, 238-249, doi:10.1016/j.str.2013.10.020 (2014).
- 47 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).
- 48 Schwede, T. Protein modeling: what happened to the "protein structure gap"? *Structure* **21**, 1531-1540, doi:10.1016/j.str.2013.08.007 (2013).
- 49 Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **44**, 98-104 (1958).
- 50 Ramakrishnan, V. The ribosome emerges from a black box. *Cell* **159**, 979-984, doi:10.1016/j.cell.2014.10.052 (2014).
- 51 Wiesmeyer, H. & Cohn, M. Purification and Characterization of Amylomaltase Induced in Escherichia-Coli. *Federation Proceedings* **16**, 270-270 (1957).
- 52 Jirgensons, B. Optical Rotation and Viscosity of Native and Denatured Proteins .10. Further Studies on Optical Rotatory Dispersion. *Archives of Biochemistry and Biophysics* **74**, 57-69, doi:Doi 10.1016/0003-9861(58)90199-1 (1958).
- 53 Bode, W., Fehlhammer, H. & Huber, R. Crystal structure of bovine trypsinogen at 1-8 Å resolution. I. Data collection, application of patterson search techniques and preliminary structural interpretation. *J Mol Biol* **106**, 325-335 (1976).
- 54 Manalan, A. S. & Klee, C. B. Activation of calcineurin by limited proteolysis. *Proc Natl Acad Sci U S A* **80**, 4291-4295 (1983).
- 55 Levintha.C. Are There Pathways for Protein Folding. *Journal De Chimie Physique Et De Physico-Chimie Biologique* **65**, 44-& (1968).

- 56 Anfinsen, C. B. Principles That Govern Folding of Protein Chains. *Science* **181**, 223-230, doi:Doi 10.1126/Science.181.4096.223 (1973).
- 57 Karplus, M. The Levinthal paradox: yesterday and today. *Folding & Design* **2**, S69-S75, doi:Doi 10.1016/S1359-0278(97)00067-9 (1997).
- 58 Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598-1603 (1991).
- 59 Fisher, C. K. & Stultz, C. M. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol* **21**, 426-431, doi:10.1016/j.sbi.2011.04.001 (2011).
- 60 Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-19 (1997).
- 61 Kumar, S., Ma, B., Tsai, C. J., Sinha, N. & Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* **9**, 10-19, doi:10.1110/ps.9.1.10 (2000).
- 62 Miller, D. W. & Dill, K. A. Ligand binding to proteins: the binding landscape model. *Protein Sci* **6**, 2166-2179, doi:10.1002/pro.5560061011 (1997).
- 63 Ma, B., Kumar, S., Tsai, C. J. & Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng* **12**, 713-720 (1999).
- 64 Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078-1082, doi:10.1038/nature06232 (2007).
- 65 Grunberg, R., Leckner, J. & Nilges, M. Complementarity of structure ensembles in protein-protein binding. *Structure* **12**, 2125-2136, doi:10.1016/j.str.2004.09.014 (2004).
- 66 Wlodarski, T. & Zagrovic, B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc Natl Acad Sci U S A* **106**, 19346-19351, doi:10.1073/pnas.0906966106 (2009).
- 67 Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **35**, 539-546, doi:10.1016/j.tibs.2010.04.009 (2010).
- 68 Clore, G. M. Interplay between conformational selection and induced fit in multidomain protein-ligand binding probed by paramagnetic relaxation enhancement. *Biophys Chem* **186**, 3-12, doi:10.1016/j.bpc.2013.08.006 (2014).
- 69 Weikl, T. R. & von Deuster, C. Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. *Proteins* **75**, 104-110, doi:10.1002/prot.22223 (2009).

- 70 Junker, J. P., Ziegler, F. & Rief, M. Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science* **323**, 633-637, doi:10.1126/science.1166191 (2009).
- 71 Ganguly, D., Zhang, W. & Chen, J. Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. *Mol Biosyst* **8**, 198-209, doi:10.1039/c1mb05156c (2012).
- 72 Kiefhaber, T., Bachmann, A. & Jensen, K. S. Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr Opin Struct Biol* **22**, 21-29, doi:10.1016/j.sbi.2011.09.010 (2012).
- 73 Zhang, W., Ganguly, D. & Chen, J. Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins. *PLoS Comput Biol* **8**, e1002353, doi:10.1371/journal.pcbi.1002353 (2012).
- 74 Gianni, S., Dogan, J. & Jemth, P. Distinguishing induced fit from conformational selection. *Biophys Chem* **189**, 33-39, doi:10.1016/j.bpc.2014.03.003 (2014).
- 75 Vogt, A. D., Pozzi, N., Chen, Z. & Di Cera, E. Essential role of conformational selection in ligand binding. *Biophys Chem* **186**, 13-21, doi:10.1016/j.bpc.2013.09.003 (2014).
- 76 Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, 2-8, doi:10.1016/j.tibs.2007.10.003 (2008).
- 77 Borg, M. *et al.* Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc Natl Acad Sci U S A* **104**, 9650-9655, doi:10.1073/pnas.0702580104 (2007).
- 78 McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977).
- 79 Mccammon, J. A. Protein Dynamics. *Reports on Progress in Physics* **47**, 1-46, doi:Doi 10.1088/0034-4885/47/1/001 (1984).
- 80 Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789-796, doi:10.1038/nchembio.232 (2009).
- 81 Boehr, D. D. & Wright, P. E. Biochemistry. How do proteins interact? *Science* **320**, 1429-1430, doi:10.1126/science.1158818 (2008).
- 82 Namanja, A. T. *et al.* Toward flexibility-activity relationships by NMR spectroscopy: dynamics of Pin1 ligands. *J Am Chem Soc* **132**, 5607-5609, doi:10.1021/ja9096779 (2010).

- 83 Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR perspective on enzyme dynamics. *Chem Rev* **106**, 3055-3079, doi:10.1021/cr050312q (2006).
- 84 Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**, 942-968, doi:10.1016/j.bbapap.2010.10.012 (2011).
- 85 Lakomek, N. A. *et al.* Residual dipolar couplings as a tool to study molecular recognition of ubiquitin. *Biochem Soc Trans* **36**, 1433-1437, doi:10.1042/BST0361433 (2008).
- 86 Tolman, J. R. & Ruan, K. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem Rev* **106**, 1720-1736, doi:10.1021/cr040429z (2006).
- 87 Loria, J. P., Berlow, R. B. & Watt, E. D. Characterization of enzyme motions by solution NMR relaxation dispersion. *Acc Chem Res* **41**, 214-221, doi:10.1021/ar700132n (2008).
- 88 Tollinger, M. *et al.* An isolated helix persists in a sparsely populated form of KIX under native conditions. *Biochemistry* **45**, 8885-8893, doi:10.1021/bi0607305 (2006).
- 89 Lipari, G. & Szabo, A. Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry* **20**, 6250-6256 (1981).
- 90 Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* **104**, 4546-4559, doi:10.1021/ja00381a009 (1982).
- 91 Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemical Society* **104**, 4559-4570, doi:10.1021/ja00381a010 (1982).
- 92 Clore, G. M. *et al.* Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins. *Journal of the American Chemical Society* **112**, 4989-4991, doi:10.1021/Ja00168a070 (1990).
- 93 Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* **16**, 23-28 (2000).
- 94 d'Auvergne, E. J. & Gooley, P. R. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *Journal of Biomolecular Nmr* **40**, 121-133, doi:10.1007/S10858-007-9213-3 (2008).

- 95 d'Auvergne, E. J. & Gooley, P. R. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *Journal of Biomolecular Nmr* **40**, 107-119, doi:Doi 10.1007/S10858-007-9214-2 (2008).
- 96 Amin, N. T. *et al.* High-resolution NMR studies of structure and dynamics of human ERp27 indicate extensive interdomain flexibility. *Biochem J* **450**, 321-332, doi:10.1042/BJ20121635 (2013).
- 97 Kim, T. H. *et al.* The role of ligands on the equilibria between functional states of a G protein-coupled receptor. *J Am Chem Soc* **135**, 9465-9474, doi:10.1021/ja404305k (2013).
- 98 Bobby, R. *et al.* Functional implications of large backbone amplitude motions of the glycoprotein 130-binding epitope of interleukin-6. *FEBS J* **281**, 2471-2483, doi:10.1111/febs.12800 (2014).
- 99 Ortega, G., Castano, D., Diercks, T. & Millet, O. Carbohydrate affinity for the glucose-galactose binding protein is regulated by allosteric domain motions. *J Am Chem Soc* **134**, 19869-19876, doi:10.1021/ja3092938 (2012).
- 100 Kunii, Y. *et al.* The immunohistochemical expression profile of osteopontin in normal human tissues using two site-specific antibodies reveals a wide distribution of positive cells and extensive expression in the central and peripheral nervous systems. *Med Mol Morphol* **42**, 155-161, doi:10.1007/s00795-009-0459-6 (2009).
- 101 Brown, L. F. *et al.* Expression and distribution of osteopontin in human tissues: widespread association with luminal epithelial surfaces. *Mol Biol Cell* **3**, 1169-1180 (1992).
- 102 Weber, G. F. The metastasis gene osteopontin: a candidate target for cancer therapy. *Biochim Biophys Acta* **1552**, 61-85 (2001).
- 103 Rittling, S. R. & Chambers, A. F. Role of osteopontin in tumour progression. *Br J Cancer* **90**, 1877-1881, doi:10.1038/sj.bjc.6601839 (2004).
- 104 Weber, G. F., Lett, G. S. & Haubein, N. C. Osteopontin is a marker for cancer aggressiveness and patient survival. *Br J Cancer* **103**, 861-869, doi:10.1038/sj.bjc.6605834 (2010).
- 105 Brown, L. F. *et al.* Osteopontin expression and distribution in human carcinomas. *Am J Pathol* **145**, 610-623 (1994).
- 106 Christensen, B., Nielsen, M. S., Haselmann, K. F., Petersen, T. E. & Sorensen, E. S. Post-translationally modified residues of native human osteopontin are located in clusters: identification of 36 phosphorylation and five O-glycosylation sites and their biological implications. *Biochem J* **390**, 285-292, doi:10.1042/BJ20050341 (2005).

- 107 Prince, C. W. *et al.* Isolation, characterization, and biosynthesis of a phosphorylated glycoprotein from rat bone. *J Biol Chem* **262**, 2900-2907 (1987).
- 108 McKee, M. D. & Nanci, A. Osteopontin: an interfacial extracellular matrix protein in mineralized tissues. *Connect Tissue Res* **35**, 197-205 (1996).
- 109 Tuck, A. B., Chambers, A. F. & Allan, A. L. Osteopontin overexpression in breast cancer: knowledge gained and possible implications for clinical management. *J Cell Biochem* **102**, 859-868, doi:10.1002/jcb.21520 (2007).
- 110 Ramchandani, D. & Weber, G. F. Interactions between Osteopontin Vascular Endothelial Growth Factor: Implications for Cancer. *Biochim Biophys Acta*, doi:10.1016/j.bbcan.2015.02.003 (2015).
- 111 Wai, P. Y. & Kuo, P. C. The role of Osteopontin in tumor metastasis. *J Surg Res* **121**, 228-241, doi:10.1016/j.jss.2004.03.028 (2004).
- 112 Rodrigues, L. R., Teixeira, J. A., Schmitt, F. L., Paulsson, M. & Lindmark-Mansson, H. The role of osteopontin in tumor progression and metastasis in breast cancer. *Cancer Epidemiol Biomarkers Prev* **16**, 1087-1097, doi:10.1158/1055-9965.EPI-06-1008 (2007).
- 113 Agnihotri, R. *et al.* Osteopontin, a novel substrate for matrix metalloproteinase-3 (stromelysin-1) and matrix metalloproteinase-7 (matrilysin). *J Biol Chem* **276**, 28261-28267, doi:10.1074/jbc.M103608200 (2001).
- 114 Platzer, G. *et al.* The metastasis-associated extracellular matrix protein osteopontin forms transient structure in ligand interaction sites. *Biochemistry* **50**, 6113-6124, doi:10.1021/bi200291e (2011).

Chapter 1

Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin

Platzer, G.,* Kurzbach, D.,* Schwarz, T.C., Henen, M.A., Konrat, R., Hinderberger, D. *Biochemistry* **52**, 5167-5175 (2013)

*both authors contributed equally

ABSTRACT

Intrinsically disordered proteins (IDPs) constitute a class of biologically active proteins that lack defined tertiary and often secondary structure. The IDP Osteopontin (OPN), a cytokine involved in metastasis of several types of cancer, is shown to simultaneously sample extended, random coil-like conformations and stable, cooperatively folded conformations. By a combination of two magnetic resonance methods, electron paramagnetic resonance (EPR) and nuclear magnetic resonance (NMR) spectroscopy, we demonstrate that the OPN ensemble exhibits not only characteristics of an extended and flexible polypeptide, as expected for an IDP, but simultaneously those of globular proteins, in particular sigmoidal structural denaturation profiles. Both types of states, extended and cooperatively folded, are populated simultaneously by OPN in its apo state. The heterogeneity of structural properties of IDPs is thus shown to even involve cooperative folding and unfolding events.

Introduction

Intrinsically disordered proteins (IDPs) have revolutionized structural biology in recent years. Despite a lack of well-folded, crystallizable structure in the conventional sense, they fulfill essential functions in eukaryotic life and thus challenge the traditional structure-function paradigm.^{1,2} The flexible and dynamic structure of IDPs and their ability to adopt different functional structures (e.g., folding upon binding) yet allow for multiple interactions of a particular protein with several binding partners.^{3,4} This makes IDPs intriguing substrates for studies in modern proteomics. Also from a biophysical and structural biology point of view, IDPs remain puzzling in many aspects. Because of the limited number of experimental techniques suited for investigations of IDPs, their solution states, conformational space and modes of conformational sampling are not well understood.^{2,5,6} However, there is a growing body of evidence that these proteins commonly classified as disordered or unstructured should be conceived as “ensembles of a continuum of rapidly interconverting structures”⁷ that contain a heterogeneous assembly of conformations, ranging from

random coils to compact structures that have regions of higher tendencies for secondary structures.⁸⁻¹⁰ Often, preformed local secondary structure elements comprise epitopes for biologically relevant protein interactions.¹¹⁻¹⁴ Although these partially preformed elements typically undergo folding-upon-binding events resulting in stable structural arrangements of separated interaction elements no distinct tertiary structure is observed in the apo state.

Motivated by the fact that in the past several years intrinsically local magnetic resonance techniques, especially nuclear magnetic resonance (NMR), have led to intriguing insight into conformational and dynamic properties of IDPs,¹⁵ we here apply solution-state NMR in combination with frozen-state electron paramagnetic resonance (EPR) spectroscopy on an IDP. We aim to combine data gained from a dynamic system state (NMR) with data about astatic snapshot of the system (EPR) to gain a detailed picture of structural transition events that are potentially comprised in the conformational space of an IDP. We have chosen Osteopontin (OPN) as a model compound, because earlier studies have shown that this IDP exhibits interesting structural properties like preformed ligand binding sites and a varying compactness profile along the disordered protein backbone. From a biological point of view OPN is a cytokine involved in metastasis of several kinds of cancer (see ref. 16 for a biophysical characterization of OPN).^{4,16} Here we show through the combination of EPR and NMR that OPN is also interesting from a biophysical point of view. The compound samples a broad distribution of compact and expanded conformations as expected for an IDP, and conformational sampling also comprises cooperative folding and unfolding events. Cooperative folding is well documented in classical proteomics, where transitions between random coil and globular states with distinct long-range interactions are typically described as first order processes. These transitions are in most cases sigmoidal in nature,¹⁷ and different conformations constitute energetically different thermodynamic states.¹⁸ The unexpected finding reported here requires that the classification of IDPs in terms of rapidly interconverting structures has to be augmented with simultaneous conformational sampling of extended as well as cooperatively folded conformations, i.e., with the fact that in IDPs different conformations of a single protein may interconvert via cooperative (phase) transitions.

Initial studies combining NMR and EPR on partially unfolded proteins have already been published.¹⁹ The complementary combination of EPR with NMR spectroscopy applied here leads to coarse-grained information about the

conformational states of the disordered OPN (EPR) as well as detailed information about its folded conformations (NMR). This combined magnetic resonance methodology and data interpretation may be applicable to other disordered protein systems.²⁰⁻²⁴

Experimental Procedures

Protein Preparation. The expression and purification of recombinant quail OPN protein (OPN220) mutants were performed as described previously.¹⁶ All details of protein expression, purification, and spin labeling are given in ref 16. Essentially, cysteine mutations were introduced using the QuickChangeII site-directed mutagenesis kit (Stratagene). For NMR and EPR analysis, all protein samples were concentrated to 0.8 mM in phosphate buffer [50 mM sodium phosphate and 50 mM NaCl (pH 6.5)] in a 90% H₂O/D₂O mixture. EPR double mutants were tagged with the nitroxide spin label (1-oxyl-2,2,5,5-tetramethyl- Δ^3 -pyrroline-3-methyl) methanethiosulfonate (MTSL), in a process analogous to the labeling procedure described in ref 16. For the purpose of this work, the choice of the label is, however, not crucially important, because the increase in ring rigidity one would gain by changing to PROXYL is negligible. The protein mutants were subject to rigorous purification using PD-10 desalting columns to remove all excess spin labels. The labeling efficiency was determined by means of 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) and UV-vis absorbance. The labeling efficiency was always >95%.

Experimental Double Electron-Electron Resonance (DEER). DEER is applied to glassy solids obtained by freeze-quenching the solutions after addition of 30% (v/v) glycerol. [Note that the presence of 20-30% (v/v) glycerol in buffered solutions of proteins is known to not affect protein conformations significantly.⁴⁴] This is achieved by immersing the sample tube in supercooled isopentane. In this way, a snapshot representative for the solution at the glass transition temperature is detected. Prior to being freeze-quenched, the samples were transferred to 3mm outer diameter quartz tubes. The sample volume was approximately 100 μ l and always large enough to fill the complete sensitive volume in the resonator. The four pulse DEER sequence $\pi/2(\nu_{\text{obs}}) - \tau_1 - \pi(\nu_{\text{obs}}) - (\tau_1 + t) - \pi(\nu_{\text{pump}}) - (\tau_2 - t) - \pi(\nu_{\text{obs}}) - \tau_2 - \text{echo}$ was used to obtain

dipolar time evolution data at X-band frequencies (9.2 to 9.4 GHz) with a Bruker Elexsys 580 spectrometer equipped with a Bruker Flexline split-ring resonator (ER4118X_MS3). The dipolar evolution time t was varied, whereas τ_2 (3 μ s) and τ_1 were kept constant. Proton modulation was averaged by the addition of eight time traces of variable τ_1 values, starting with $\tau_{1,0} = 200$ ns and using increments ($\Delta\tau_1$) of 8 ns. The resonator was overcoupled to $Q \approx 100$. The pump frequency, ν_{pump} , was set to the maximum of the EPR spectrum. The observer frequency, ν_{obs} , was set to $\nu_{\text{pump}} + 61.6$ MHz, coinciding with the low-field local maximum of the nitroxide spectrum. The observer pulse lengths were 32 ns for both $\pi/2$ and π pulses, and the pump pulse length was 12 ns. The temperature was set to 50 K by cooling with a closed cycle cryostat (ARS AF204, customized for pulse EPR, ARS, Macungie, PA). The total measurement time for each sample was ~ 12 h. The resulting time traces were normalized to $t = 0$.

DEER Background Correction. Background correction was conducted by dividing by experimental functions gained by DEER on four spin labeled single mutants (C54, C108, C188, C427). All single-mutant data corresponded to a homogeneous three-dimensional distribution of spins (homogeneous exponential decay). Measurements on single mutants were taken at similar protein concentrations as in the case of the double mutants (0.8 mM) to ensure a similar excluded volume. Because the spin concentration was therefore only half the concentration as in the case of double mutants, this adds a factor of 2 to the exponent of the homogeneous background function to compensate for the lower concentration:

$$V_{\text{DEER}}^{\text{hom}} = \exp\left(-\frac{2\pi\mu_0 g_1 g_2 \beta_e}{9\sqrt{3}\hbar} \lambda c_{\text{pump}} t^{d/3}\right) \quad (1)$$

where c_{pump} is the concentration of the spins resonant at frequency ν_{pump} and $d = 3$. g_1 and g_2 are the g values of the observer and the pump spin, respectively, in a two-spin system. μ_0 is the magnetic moment of the electron spin. β_e is the Bohr Magneton. λ is an experimental constant that denotes the fraction of spins flipped by the pump pulse ($\hbar = h/2\pi$). The concentration factor of 2 can, however, be neglected (as we did for background correction) for the relative comparison used for analysis of the data presented here. *Determination of Δ_{eff} .* Δ_{eff} was determined by fitting the background-corrected DEER data with a Gaussian distance distribution according to

eqs S3–S5 of the Supporting Information. As such, a $V_R(t)$ was determined from the fit and $\Delta_{\text{eff}} = 1 - V_R(t = 3 \mu\text{s})$.

Experimental NMR Spectroscopy. NMR spectra were recorded at 20°C on Varian spectrometers operating at 500, 600 and 800 MHz. OPN220 protein samples were dissolved in 50mM sodium phosphate and 50mM NaCl (pH 6.5) with 10% D₂O as lock solvent. PRE intensity ratios were derived from pulsed field gradient (PFG) sensitivity-enhanced two-dimensional ¹H-¹⁵N HSQC spectra of ¹⁵N labeled OPN220 mutants C54, C108, C188 and C247 respectively.⁴⁵ NMR spectra were processed using NMRPipe⁴⁶ and analyzed using SPARKY.

Paramagnetic Relaxation Enhancements (PREs). Single cysteine OPN220 mutants (C54, C108, C188 and C247) were tagged with the nitroxide spin label (1-oxy-2,2,5,5-tetramethyl-Δ³-pyrroline-3-methyl) methanethiosulfonate (MTSL). The overall PRE effect on OPN220 was measured as the intensity ratio of cross-peaks in presence ($I_{+\text{MTSL}}$) and absence ($I_{-\text{MTSL}}$) of the cysteine-attached spin labels, as $\Delta_{\text{MTSL}} = I_{+\text{MTSL}}/I_{-\text{MTSL}}$. To account for possible interactions of the spin-label with the protein, we added MTSL to untagged, ¹⁵N-labeled OPN220 at a final concentration of 1mM. The intensity ratios (peak intensity) of OPN220 with unbound and free ($I_{+\text{FREEMTSL}}$) and without ($I_{-\text{MTSL}}$) MTSL $\Delta_{\text{UNSPEC}} = I_{+\text{FREEMTSL}}/I_{-\text{MTSL}}$ were combined with the intensity ratios from attached MTSL (Δ_{MTSL}) to calculate the PRE effect on the protein: $\text{PRE} = \Delta_{\text{MTSL}} + (1 - \Delta_{\text{UNSPEC}})$. ΔPRE was calculated as the difference in signal heights under nondenaturing conditions and high-urea and -salt conditions as $\Delta\text{PRE} = {}^{15}\text{N}-{}^1\text{H}$ HSQC intensity (high urea and salt) – ¹⁵N-¹H HSQC intensity (no urea or salt). The usage of two different samples in the presence and absence of free MTSL is necessary to ensure that MTSL itself does not have an intrinsic binding affinity for certain preferential protein sites. This cannot be probed by reduction of covalently attached MTSL, because the native protein conformations could not be separated from MTSL-induced folding in this case.

It should be noted that significant spectral overlap at 8M urea and high NaCl concentrations precluded the extraction of a complete PRE data set. Additionally, PRE measurements under complete denaturation conditions using both high urea and NaCl concentrations were not possible because of substantial viscosity effects (and thus substantially broadened NMR signals).

¹³C-¹H HSQC NMR. Reductive methylation procedures were performed as described by Means and Feeney.⁴⁷ The protein was dialyzed against 10 mM HEPES,

150 mM NaCl, and 1mM DTT buffer (pH 7.4.); 0.25 ml of 1.6 mM borane dimethylamine complex $[(\text{CH}_3)_2\text{NH}\cdot\text{BH}_3]$ and 0.5 ml of 1.6 mM ^{13}C -labeled formaldehyde were added to 0.5 ml of 0.1 mM protein, and the reaction was incubated while stirring at 4°C. Subsequently, the addition of the borane ammonium complex and ^{13}C formaldehyde was repeated and the reaction was incubated for additional 2 h. After adding 0.12 ml of 1.6 mM borane ammonium complex solution had been added, the reaction mixture was incubated at 4°C while stirring overnight. The reaction was quenched by adding glycine to yield a concentration of 200 mM. Undesired reaction products as well as excess reagents were removed by dialysis against Tris-buffer (pH 7.4). The sample was concentrated to a final concentration of approximately 0.1 mM. Two-dimensional ^{13}C - ^1H HSQC NMR experiments were conducted with synthesized ^{13}C -methylated Osteopontin on a Varian Innova 600 MHz spectrometer.

Results and Discussion

We first probe structural preferences of OPN by applying the EPR-based method double electron-electron resonance (DEER) spectroscopy to six spin-labeled Cys-double mutants of 220 residues long truncated OPN [residues 45-264 of the native protein; we denote the spin-labeled OPN double mutants as Cx-Cy (n/3 E, L, or S) with x and y being the labeling sites, n/3 the fraction spanned by the respective mutant and E,L or S the shape (exponential, linear, or sigmoidal, respectively) of the denaturation profile as will be explained below]. Mutants C54-C108 (1/3 L), C108-C188 (1/3 L) and C188-C247 (1/3 E) each span approximately 1/3 of the whole protein, the mutants C54-C188 (2/3 S) and C108-C247 (2/3 L) about 2/3 and the mutant C54-C247 (3/3 S) nearly the whole truncation mutant (see Figure 1 for a schematic representation of the spanned ranges). Conformational stabilities, understood here as resistance to urea unfolding, of these individual structural segments of OPN are investigated by recording DEER time traces for the different double mutants in dependence of increasing urea concentration. In a second step, we investigate cooperatively compacted states by means of paramagnetic relaxation enhancements and rationalize our results on basis of noncovalent structuring principles in OPN.

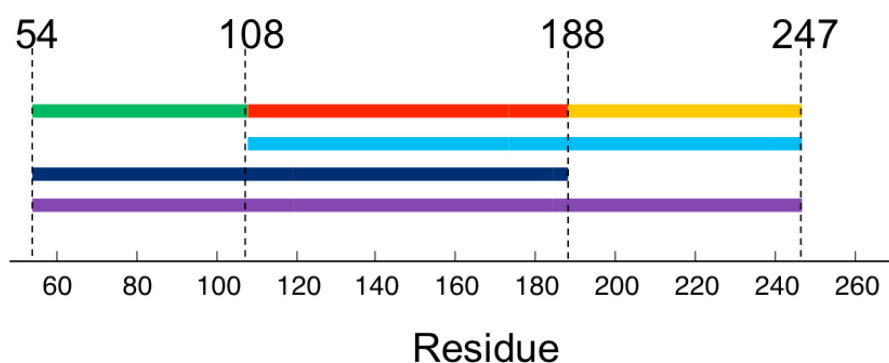


Figure 1. Scheme of spin labeling sites along the protein backbone and sketch of residues spanned by each of the six OPN double mutants.

Cooperative Transition Events. DEER experiments yield nonaveraged data [i.e., the superposition of data from every single OPN molecule in the shock-frozen solution (see Experimental Double Electron-Electron Resonance (DEER) and section 1 and Figure S1 for a detailed description of DEER)] that, after background correction that eliminates intermolecular contributions, display intramolecular dipole-dipole couplings between the two spins of the labels of double mutants as damped cosine modulation of time domain traces. These traces are the evolution of echo intensity with interpulse delay for the four pulse DEER sequence as described in the Experimental section.²²

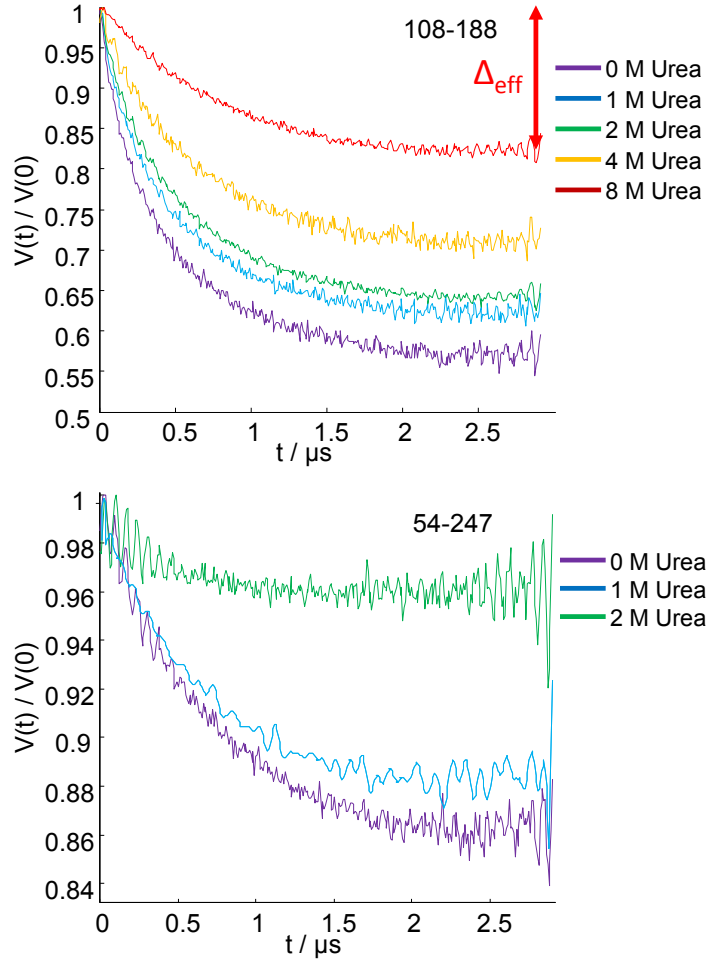


Figure 2. DEER time traces of C108-C188 (1/3 L) and C54-C247 (3/3 S) at different urea concentrations. Δ_{eff} is defined as the signal decay at $t = 3 \mu\text{s}$, as indicated by the double-headed arrow. (Note the different $V(t)/V(0)$ scales.)

In particular the modulation is related to the dipolar coupling frequency that in turn depends on the interspin distance, R , as R^{-3} .²² Typical DEER data for an OPN double mutant at different urea concentrations are shown in Figure 2 for two different double mutants [C108-C188 (1/3 L), representative for a double mutant defining a small OPN segment, and C54-C247 (3/3 S), representative for a large segment]. No clear-cut modulations are observable after experimental background correction. This indicates that the pair-distribution functions, $P(R)$, between the two spin labels of the two double mutants (R being the distance between the two spins) are quite broad, because the sum over varying damped cosines converges to exponential decay functions. This is expected for an IDP with very broad conformational ensembles like OPN,¹⁶ because every conformation (corresponding to a certain R) gives rise to a

certain modulation function (the complete data set for all double mutants is shown in the Figures S2 and S3 of the Supporting Information; all time traces are devoid of modulations).²⁵ In contrast, globular proteins frequently display DEER time traces with significant and apparent dipolar modulations because their narrow conformational ensembles give rise to a restricted interspin distance range (see Figure S1b of the Supporting Information for time traces calculated from a single discrete interspin distance).²¹ Because the established standard analysis methods²⁶ cannot be applied for the nonmodulated DEER data under investigation, we analyze DEER time traces through an effective modulation depth, Δ_{eff} (as sketched in Figure 2 and Figure S1b of the Supporting Information), which denotes the total signal decay at a t_{max} of 3 μs . Such, $\Delta_{\text{eff}} = 1 - V(t = 3 \mu\text{s})/V(t = 0)$. $V(t)$ is the DEER echo intensity at time t [for details on the determination of Δ_{eff} , see Experimental Double Electron-Electron Resonance (DEER)]. Three microseconds is with our setup the longest achievable experimental DEER evolution time for this study but is generally arbitrary for the proposed analysis of very broad distance distributions reflected in large conformational ensembles. Δ_{eff} is an approximate measure of the average interspin distance for broad $P(R)$ s. For broad distance distributions Δ_{eff} decreases with increasing interspin distance R . As such, a decrease of Δ_{eff} with increasing urea concentration is representative for unfolding and expansion of a doubly spin labeled protein of interest. This is shown and explained in detail in section 1 of the Supporting Information and graphically illustrated in Figure S1b,c of the Supporting Information for calculated data.

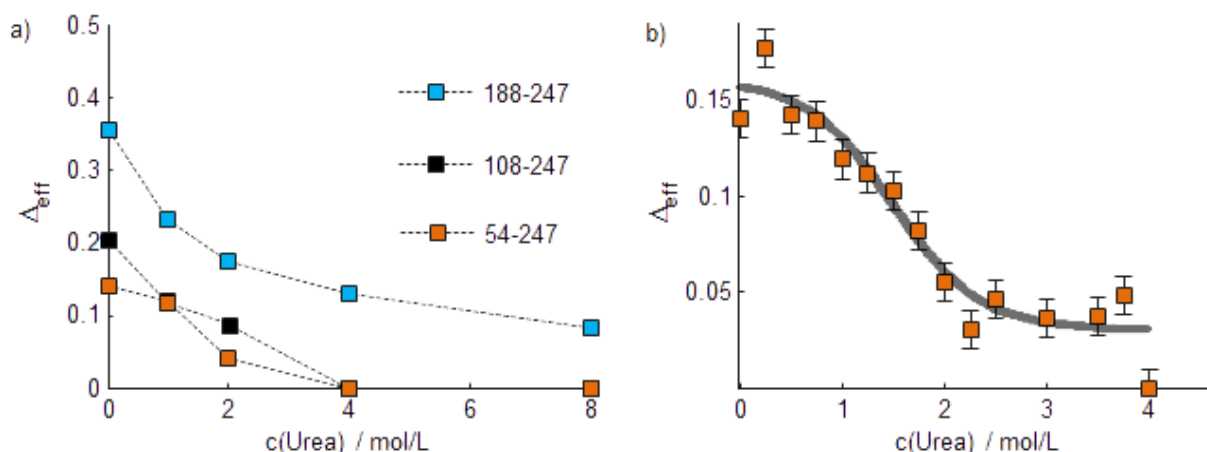


Figure 3. (a) Δ_{eff} for selected double mutants as a function of urea concentration. All the data for all double mutants under investigation can be found in Figures S2-S4 of the Supporting Information. (b) Detailed representation of Δ_{eff} for C54-C247 (3/3 S) as a function of urea concentration. Error bars stem from signal noise. The gray curve is based on a sigmoidal data fit to confirm the visual observation of sigmoidality. The fit is based on the relationship $\Delta_{\text{eff}} = a + b / \{1 + \exp[-c(\text{urea}) - m]/s\}$.

In Figure 3a, experimental Δ_{eff} values are shown as a function of urea concentration for selected double mutants comprising segments of OPN of different length (i.e., starting from the C-terminus approximately one-third, two-thirds, and the entire truncation mutant; for the entire data set, see Figures S2-S4 of the Supporting Information). For the C-terminal part of OPN, comprised by the mutant C188-C247 (1/3 E), an exponential decay of Δ_{eff} (i.e., an increase in interspin distance) can be observed with increasing urea concentration. This mutant gives rise to the steepest observed slope of any of the Δ_{eff} functions and can hence be regarded as a relative reference for the effect of conformational denaturation on unstably folded protein segments of potentially random-coil like character. Already for low urea concentrations such segments show significant conformational expansion (i.e., a decrease in Δ_{eff}) in accordance with the idea of very low stability of transient or residual structural elements in IDPs. For mutant C108-C247 (2/3 L) [as well as for mutants C54-C108 (1/3 L) and C108-C188 (1/3 L) (see Figure S4 in the Supporting Information)], one observes an approximately linear decrease of Δ_{eff} with urea concentration, indicating that the OPN segment framed by this mutant (approximately two-thirds of the protein) is on average conformationally more stable than the segment

between C188 and C247 [C188-C247 (1/3 E)], although still largely unstructured, random coil-like or (pre)molten globule-like. Strikingly, however, for the mutant C54-C247 (3/3 S) of OPN and C54-C188 (2/3 S) we observe a sigmoidal development of the Δ_{eff} -derived denaturation profiles with urea concentration (see Figure 3b). Sigmoidality is a hallmark for cooperative folding of protein conformations and unexpected for an IDP.¹⁷ The sigmoidal development of Δ_{eff} is depicted in more detail for C54-C247 (3/3 S) in Figure 3b. In Figure 4, the six spin label pairs probed through the six double mutants of OPN are sketched and labeled with the shape of the respective denaturation or Δ_{eff} profile, i.e., linear (L), exponential (E) or sigmoidal (S).

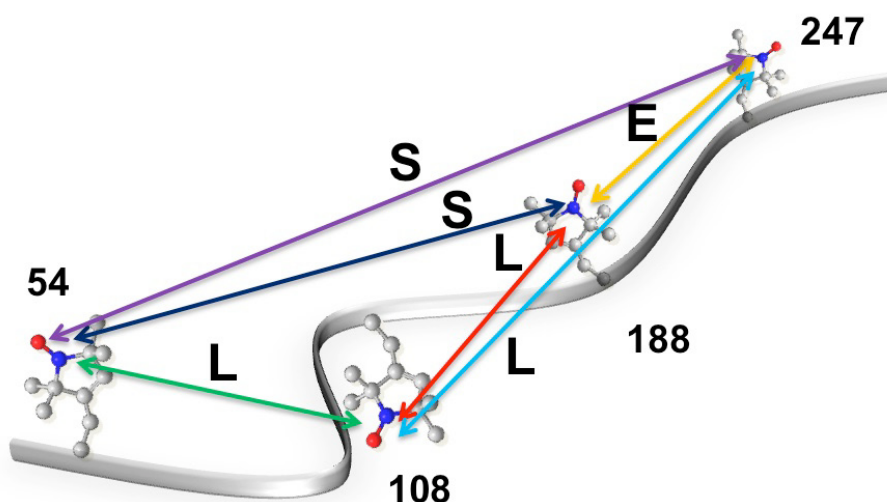


Figure 4. Sketch of OPN double mutants assessed by DEER. Labels E (exponential), L (linear) and S (sigmoidal) denote the profile shapes of the respective Δ_{eff} functions (see Figure 3 and Figures S2 and S4 of the Supporting Information).

A sigmoidal denaturation profile is indicative for stably and cooperatively folded tertiary structures of OPN, because for low urea concentrations of ≤ 0.75 M the whole protein does not expand significantly (as seen in nearly constant Δ_{eff}). This observation of a cooperatively folded conformation is surprising as $P(R)$ values for OPN are generally quite broad, which can be deduced from prior studies concerning OPN's conformational space¹⁶ and is reflected in the nonmodulated DEER time traces (see section 1 and Figures S2-S4 of the Supporting Information and Figure 2). This interesting finding can, however, be understood by concluding that the structural ensemble of OPN contains both cooperatively folded and unfolded conformations and that both contribute to the DEER signals. It should be noted that the interpretation of DEER data is complicated by the fact that for rather small separations of spin label

sites both compact and extended (sub)structures contribute significantly to the observed Δ_{eff} values. This means that compact conformations contribute more strongly to DEER time domain data of systems with distant spin labels (e.g., C54-C247 (3/3 S)). In contrast, if the spin labels are closer along the primary sequence and the mean distance becomes shorter (e.g. C188-C247(1/3 L)), longer distances dominate the DEER data. This is shown in detail in section 2 of the Supporting Information. Because cooperatively folded conformations are more compact than unfolded ones, sigmoidal Δ_{eff} profiles can therefore only be observed for double-mutants with labeling sites that are separated by more than 130 residues, because in these cases the mean distance of the conformational ensemble is large and hence the time domain data are dominated by contributions of folded conformations comprising short (electron-electron) distances. Hence, the corresponding Δ_{eff} profiles are sigmoidal. In contrast, for the three mutants that comprise only one-third of OPN, the DEER signal is dominated by contributions of extended structures. These exhibit linear or exponential denaturation profiles, lacking any sigmoidal contributions to the development of Δ_{eff} . In general one can thus state for broad $P(R)$ values that with increasing separations between two labeling sites the relative contributions of compact conformations to the Δ_{eff} profiles increase. Compact conformations consequently dominate the urea dependence of Δ_{eff} for C54-C247 (3/3 S), while extended conformations contribute more significantly to Δ_{eff} for C188-C247 (1/3 E), C108-C188 (1/3 L), and C54-C108 (1/3 L). In summary, it is important to note that only the simultaneous sampling of both extended and compact (cooperatively folded) substates in OPN leads to superposition of DEER data with significantly different urea dependence: a sigmoidal Δ_{eff} profile for the mutant C54-C247 (3/3 S) and only gradual changes for double-mutants spanning smaller segments than this mutant. Partial structuring as an underlying reason for this observation can be ruled out. For C54-C247 (3/3 S) (nearly the whole length of the truncation mutant), cooperative unfolding can be observed, while this is not the case for the inner segments comprising only approximately one-third of OPN. The latter would, however, necessarily be the case if any segment of OPN would be statically, partially structured. This deduction is possible here only because EPR of freeze-quenched solutions elucidates the whole set of coexisting conformations; ensemble averaged data here would not allow one to discern between partial structuring and sampling of compact conformations. In summary, OPN's

structural behavior comprises cooperative phase transitions events between compact and expanded conformations.

Noncovalent Structuring Principles of OPN Conformations as Seen in DEER Data. Given the enrichment (compared to the whole proteome) of polar and charged amino acids in IDPs, one can expect the stabilization of cooperative folded structures of OPN (as a necessary consequence of the cooperative phase transitions events) likely to be triggered by electrostatic interactions.³ In Figure 5 the effect of 4 M NaCl on Δ_{eff} is shown for the six double mutants under investigation in the presence and the absence of 8 M urea. Figure 5a shows Δ_{eff} values for denaturation conditions 4 M NaCl, 8 M urea, or 4 M NaCl with 8 M urea. Figure 5b shows exemplary DEER time traces for the double mutant C108-C188 under these conditions and its native state. Note that NaCl does not significantly affect the effective modulation depth in the absence of urea but does in its presence. For C108-C188 (1/3 L) and C54-C188 (2/3 S), Δ_{eff} decreases with increasing NaCl concentration even at 8 M urea. As such, the interspin distance still increases because of the increasing NaCl concentration even if 8 M urea is already present in the solution. Thus, one can state that urea alone does not expand OPN as strongly as urea in combination with NaCl. NaCl alone only has small effects on Δ_{eff} . This can be rationalized as follows. NaCl screens electrostatic interactions, while urea does not.²⁷⁻²⁹ Hence, screening of electrostatic interactions seems not to be enough to significantly expand OPN's conformations. Only complementary screening of hydrophobic interactions and hydrogen bonds by urea and screening of electrostatics through NaCl lead to the most effective expansion of OPN's conformations. Hence, one might speculate that urea alone is not sufficient to completely denature OPN and eliminate all residual structural elements from its conformations; only the combination of complementary screening agents might be sufficiently strong. This is remarkable because earlier biophysical characterizations of OPN undoubtedly classify this protein as intrinsically disordered.¹⁶ The significant electrostatic contribution to the energetics of OPN's conformational sampling modes is discussed in more detail below, taking into account paramagnetic relaxation enhancement (PRE) data. Protein dimerization as a possible source of error is ruled out through DEER measurements performed on the corresponding four single mutants (see section 3 and Figure S7 of the Supporting Information). There, completely

homogeneous spin distributions are observed, indicating that OPN does not show any form of aggregation at the concentrations (0.8 mM) used for the DEER (and NMR) measurements.

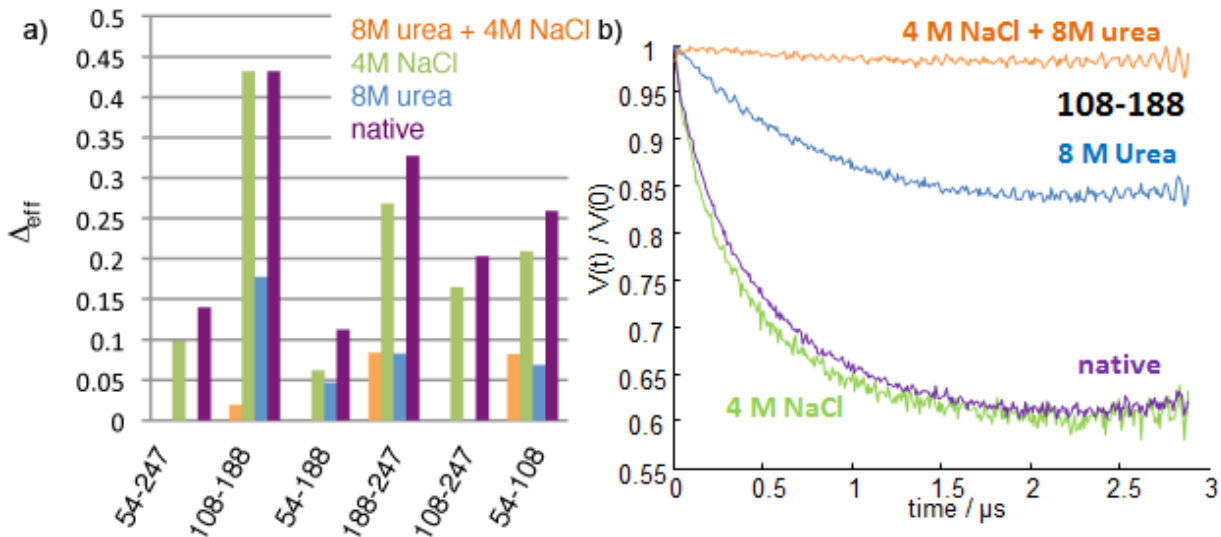


Figure 5. (a) Δ_{eff} for the different double mutants under different denaturing conditions (4 M NaCl, 8 M urea, or 4 M NaCl with 8 M urea). (b) Exemplary (for C108-C188 (1/3 L)) DEER time traces for different denaturation conditions (4 M NaCl, 8 M urea, or 4 M NaCl with 8 M urea).

It should be noted that there is a growing body of evidence for the so-called direct mechanism of urea denaturation to describe protein-urea interaction correctly.^{30,31} This mechanism states that urea directly binds to the protein backbone likely (primarily) through dispersive interactions and thereby interrupts protein structure-stabilizing interactions.^{30,31} For this case of urea denaturation of OPN, this direct mechanism might be important for gaining a full understanding of the DEER data, because the Δ_{eff} values for some cases indicate mean distances between two labeling sites that are longer than the distance one would expect in a random coil polypeptide. For example, for the mutants 54-108 one would expect distances between the two labels of ~10 nm from Δ_{eff} (see Figure S1c of the Supporting Information), while for a true random coil with a Flory characteristic ratio of 2 distances of ~6 nm would be expected.³² This discrepancy might be traced back binding of urea to the protein backbone, which leads to longer persistence lengths or rather scaling exponents and thus also to longer inter-residue distances.³³

Noncovalent Structuring Principles of OPN Conformations as Seen in NMR Data. The urea dependence of NMR backbone chemical shift ^{15}N - ^1H (cs) data was analyzed for residues of the core region and of the terminal region of OPN (see Figure 6a, Experimental NMR Spectroscopy, and Figure S5 of the Supporting Information). The data show only marginal chemical shift changes (Δcs) observed below 1 M for some residues in the compact core region (171 and to some degree 144), and Δcs increases substantially only with 2 M urea or more. For most residues in and outside the core region a more or less steady increase in Δcs can be observed. The core regions is approximately located between residues 100 and 180 (see Figure S5 for more cs data for residues of the core segment of OPN).¹⁵ Overall, larger chemical shift changes were observed for residues located in the compact core of OPN (100-180). Most importantly, slight deviations from the linear Δcs versus urea concentration behavior was observed for residues in the core segment (171 and 144). It should be noted that all conformational substates of OPN contribute to the observed chemical shift changes. Given the small population of the compact structure only small contributions can be expected. Although the cooperative phase transition observed by means of EPR, indicating the existence of rather stable conformations of OPN that resist denaturation by smaller urea concentrations, cannot generally be reproduced through Δcs , the chemical shift changes clearly provide additional evidence for a more compact segment in OPN located between residues 100 and 180.^{4,16} As such, the Δcs data are not in conflict with the interpretation derived above from DEER.

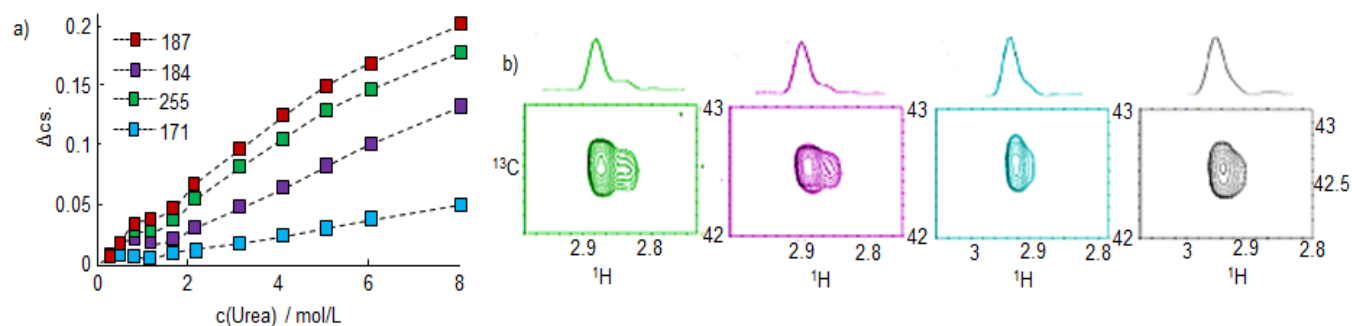


Figure 6. (a) ^{15}N - ^1H NMR chemical shift changes {calculated as $\text{cs}[c(\text{urea}) = x \text{ M}] - \text{cs}[c(\text{urea}) = 0 \text{ M}]$ } of selected backbone positions as a function of urea concentration. (b) ^{13}C - ^1H HSQC of $^{13}\text{CH}_3$ -Lys-labeled side chains: green for 0 M urea, pink for 2 M urea, blue for 4 M urea, and black for 6 M urea. Note that the shift in the ^1H dimension is merely a consequence of readjusting the transmitter offset in dependence of the urea concentration to achieve suppression of water signals.

An additional indication for the existence of compact structures to the conformational ensemble was provided by NMR observations of side chain positions. In Figure 6b, data from ^{13}C - ^1H HSQC [heteronuclear single-quantum coherence (see experimental ^{13}C - ^1H HSQC NMR)] of $^{13}\text{CH}_3$ -Lys-labeled OPN are shown. The majority of cross peaks is overlapped and stems from side chains of residues in random-coil like conformations (signal at higher number of parts per million of the ^1H dimension), which are typically more solvent exposed and flexible than residues in folded protein segments. However, a fraction of methyl cross peaks (approximately 20% as determined from fitting signal volumes; signal at lower parts per million of the ^1H dimension) is significantly shifted from the bulk signals. This shows that a fraction of the lysine residues in the conformational ensemble are exposed to an environment that is different from that observed for random coil polypeptides. This might indicate that the conformational ensemble partially exists in a compacted form in which the lysine residues are embedded in a more water-depleted core. As such, the ^{13}C - ^1H HSQC is not in conflict with the EPR-derived conclusion that part of the conformational ensemble of OPN cooperatively folds into compact conformations. The shifted lysine peaks even remain unchanged and clearly separated from the bulk of lysine side chains at urea concentrations of $\leq 1 \text{ M}$. We refrain here from analyzing this observation in the context of cooperativity, yet it further supports the existence of compact structures in the ensemble of OPN. The NMR ^{15}N - ^1H chemical shift changes

and ^{13}C - ^1H HSQC data (that is, backbone- as well as side chain-based data) are in agreement with a compact, presumably cooperatively folded substate in the conformational ensemble of OPN besides large fractions of extended conformations.

Paramagnetic Relaxation Enhancements. Because in a hypothetical random coil polymer paramagnetic relaxation enhancements (PRE; i.e., enhanced relaxation rates of nuclear resonances due to the presence of an electron spin) are limited to residues flanking the spin label sites, the observation of specific and sizeable long-range PRE effects provides unambiguous evidence for the existence of compact states.^{34,35} PRE effects were measured for the four single mutants C54, C108, C188 and C247. The different PRE-residue plots (see Figure 7a) show that the conformational ensemble of OPN indeed features distinct long-range interactions. Specifically, the PRE results obtained for the mutants C108 and C188 provide clear evidence for the prevalence of a structurally compact region in OPN encompassing residues 100-200 (recall that intermolecular contributions can be ruled out by DEER on the single mutants at the given concentration). The structural stability of this compact conformation as a function of urea and NaCl concentration was monitored further by condition-dependent PRE changes. Figure 7b shows experimental PRE differences (ΔPRE) measured under NaCl and high urea conditions.

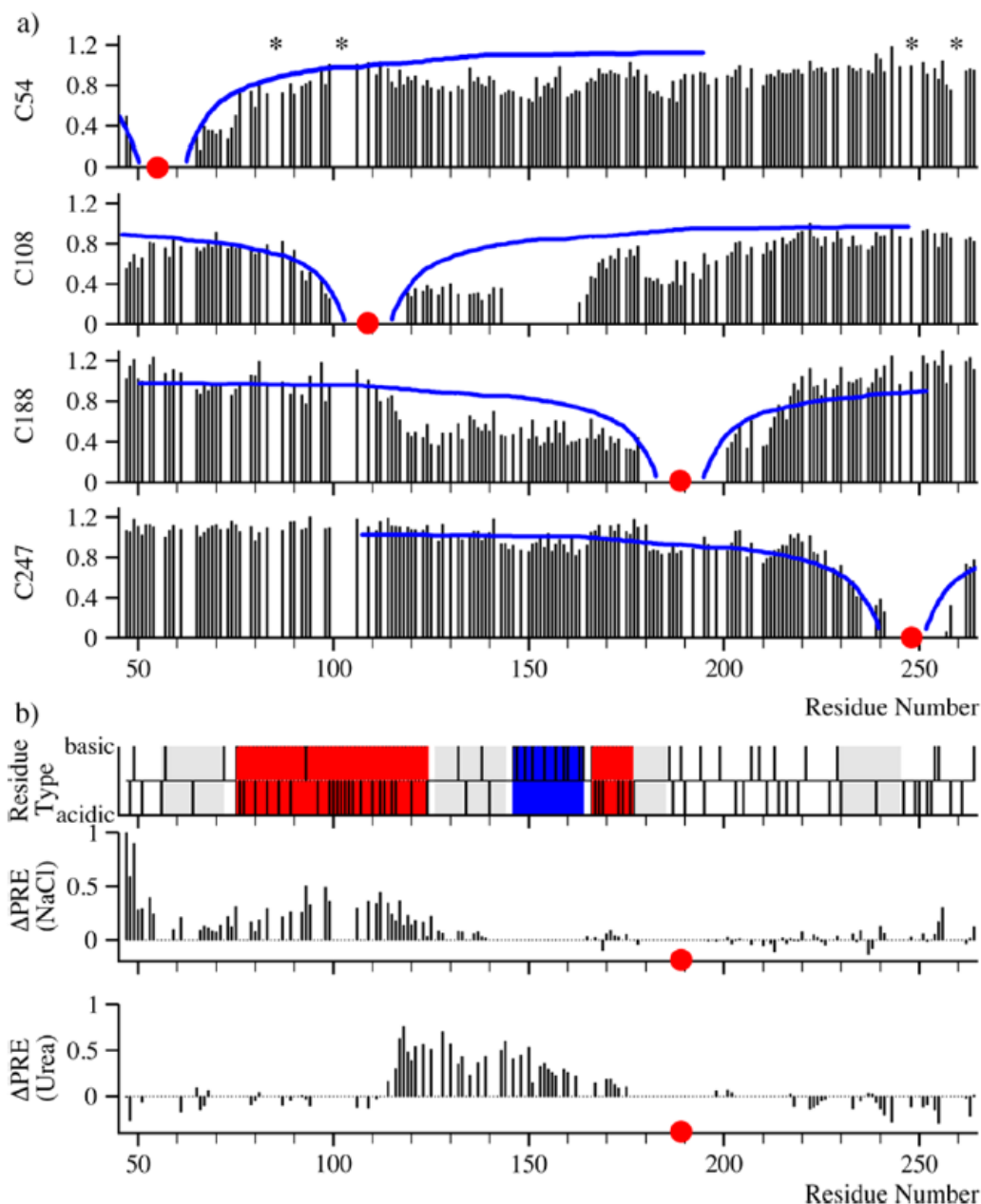


Figure 7. (a) PRE data for the four single mutants C54, C108, C188 and C247. Superimposed in blue are PREs calculated for random coils with a Flory characteristic ratio of 2 by the Solomon-Bloembergen relation.²³ The red dots mark the different labeling sites. The asterisks mark stretches comprising larger numbers of unassigned resonances. (b) A charge map of OPN (top; blue corresponds to patches of primarily basic residues, red to patches of acidic residues, and gray to primarily hydrophobic patches) and PRE changes (Δ PRE) for high salt (center) and high-urea (bottom) conditions obtained for the C188 mutant. [Δ PRE = ^{15}N - ^1H HSQC-Intensity (high urea and salt) – ^{15}N - ^1H HSQC-Intensity (no urea or salt)].

The significantly charged region encompassing residues 75-125 is nearly unaffected by the addition of urea but displays sizeable PRE changes under high NaCl conditions, while residues 125-150 are strongly effected by urea. Hence, from these results (also compare with Figure 5), we can conclude that hydrophobic interactions contribute to the structural stability of OPN and electrostatics play a pivotal role in stabilizing the compact sub-states of OPN in solution. These findings can be rationalized by a closer inspection of the charge map of OPN (Figure 7b, top). In OPN, negative charges (acidic residues, red) are concentrated in the region between residues 75 and 125, while there is a high density of positive charges (basic residues, blue) in the region between residues 145 and 165. The attraction between these positively and negatively charged regions and the hydrophobic patches around residues 60, 130 and 180 therefore suggest stabilizing interactions and consequently stronger tertiary structure propensity between residues 60 and 180, compared to other regions of OPN.¹⁶ In Figure 8, a sketch of OPN's compact conformation is shown, as one would derive it from the PRE data in Figure 7a. Long-range intrachain contacts between stretches between residue 100 and 180 as well as slight sampling of the more central regions of OPN by the two termini are depicted. In conclusion, the NMR data indicate significantly populated compact structures in OPN that are stabilized (even cooperatively stabilized as evidenced by DEER data above) by both hydrophobic and electrostatic interactions. This is also in excellent agreement with the NaCl dependence of Δ_{eff} (see Figure 5). The significant resistance to both, urea and NaCl unfolding is clearly remarkable for an IDP. It is thus reasonable to conclude that the subtle interplay between conformation-stabilizing enthalpic contributions and destabilizing entropic contributions ultimately account for OPN's conformational flexibility and its property to cooperatively sample both unfolded and cooperatively folded structures.

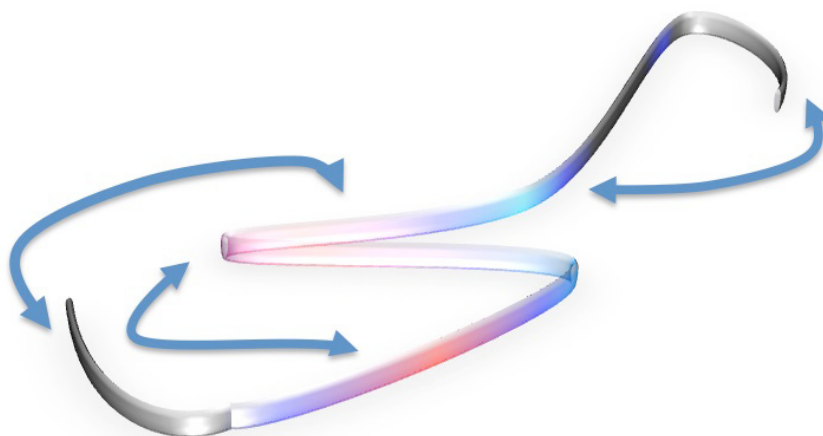


Figure 8. Sketch of the assumed “average” structure of OPN based to the PRE data. The arrows indicate significant PRE effects. As such, OPN can be pictured as having a more compact core and back-folded termini. The colors refer to the charge map in Figure 7b (blue corresponds to patches of primarily basic residues, red to patches of acidic residues, and gray to primarily hydrophobic patches).

Conclusion

Altogether, we have shown that the IDP OPN cannot be described by polymer physical models such as random coil or molten globule polymers.¹⁸ Instead, OPN simultaneously populates extended as well as cooperatively folded structures and sigmoidal molecular interconversion. This observation for OPN is a convincing experimental demonstration of conformational sampling of different thermodynamic states in an IDP.³⁶ The fact that OPN samples cooperatively stabilized overall as well as extended conformations further is particularly intriguing in the context of IDP binding mechanisms. Often protein recognition by IDPs proceeds via folding-upon-binding events accompanied by disorder-to-order transitions,³⁷ although even in the bound state IDPs (can) retain substantial conformational flexibilities (“fuzziness”),³⁸ be it static (multiple conformations) or dynamic disorder (fluctuation between different states). Protein-protein interaction is typically described either as induced fit^{39,40} or conformational selection fit.⁴⁰ While the induced fit model postulates the formation of an encounter complex followed by structural adaptation, conformational selection indicates the existence of a conformational ensemble in which the final bound state is

partly present and populated by stabilizing intermolecular interactions, although there is evidence that both mechanisms can be active simultaneously.⁴¹ The existence of a cooperatively folded substate in the structural ensemble of OPN suggests protein-protein interactions occur largely via conformational selection characterized by a significant reduction of the entropic penalty and presumably reduced fuzziness in the bound state.^{37,42} The unexpected long-range preformation of the apo state of OPN might thus be of relevance for providing specific interaction interfaces across cellular surfaces and thus endows OPN with unique possibilities to modulate interaction patterns with its several natural ligands.⁴

Furthermore, our results substantiate recent insights that urea-unfolded states of proteins differ significantly from the native state of intrinsically disordered proteins.⁴³ We show that urea can induce drastic structural rearrangements of IDPs and changes in their conformational space. This is valid in terms of elongation of end-to-end distances of random coils through coordination of urea to the protein backbone, as stated above, and secondary structure propensities become significantly altered and populations of compact substates change when urea interacts with IDP backbones directly.^{30,31} Most importantly, the existence of structural cooperative transitions from folded to unfolded states and vice versa in IDPs calls for a novel conceptual view of IDPs that goes beyond the traditional binary scheme of order versus disorder. The subtleties of heterogeneous conformational sampling in IDPs and their putative relevance for biological functions have to be adequately addressed.

Acknowledgement

We thank Christian Bauer for technical support and Prof. Hans W. Spiess for continuing support. M.A.H. acknowledges the support of Professor Dr. Hassan Eisa, Dr. Alaa El-Din Barghash and Dr. Laila A. Abouzeid (Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University). The expression plasmid for wild-type OPN was provided by the group of K. Bister (University of Innsbruck, Innsbruck, Austria).

Funding

This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG) under grant number HI 1094/2-1, the Max Planck Graduate Center with the University of Mainz (MPGC) and the Austrian Science Fund (FWF) under grant numbers P20549-N19, P22125-B12, and W1221. T.C.S. was in part funded by Forschungsstipendium 2011. M.A.H. was supported by a grant from the ÖAD entitled "KU mit Ägypten art 9/2". D.K. acknowledges support by the Gutenberg Academy of the University of Mainz.

References

- (1) Dunker, A. K.; Obradovic, Z. The protein trinity - linking function and disorder; *Nature Biotech.* **2001**, *19*, 805-806.
- (2) Tompa, P. Intrinsically unstructured proteins; *Trends Biochem. Sci.* **2002**, *27*, 527-533.
- (3) Uversky, V. N. Intrinsically disordered proteins from A to Z; *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1090-1103.
- (4) Bellahcène, A.; Castronovo, V.; Ogbureke, K. U. E.; Fisher, L. W.; Fedarko, N. S. Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer; *Nat. Rev.* **2008**, *8*, 212-226.
- (5) Konrat, R. The meandering of disordered proteins in conformational space; *Structure* **2010**, *18*, 416-449.
- (6) Uversky, V. N.; Dunker, A. K. Understanding protein non-folding; *Biochim. Biophys. Acta* **2010**, *1804*, 1231-1264.
- (7) Ozenne, V.; Schneider, R.; Yao, M.; Huang, J.-r.; Salmon, L.; Zweckstetter, M.; Jensen, M. R.; Blackledge, M. Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution; *J. Am. Chem. Soc.* **2012**, *134*, 15138-15148.
- (8) Huang, J.-r.; Grzesiek, S. Ensemble Calculations of Unstructured Proteins Constrained by RDC and PRE Data: A Case Study of Urea-Denatured Ubiquitin; *J. Am. Chem. Soc.* **2010**, *132*, 694-705.
- (9) Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with Cdc4 Subunit of an SCF Ubiquitin Ligase; *Structure* **2010**, *18*, 494-506.

- (10) Marsh, J. A.; Forman-Kay, J. D. Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints; *J. Mol. Biol.* **2009**, *391*, 359-374.
- (11) Vacic, V.; Oldfield, C. J.; Mohan, A.; Radivojac, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. Characterization of molecular recognition features, MoRFs, and their binding partners; *J. Proteome. Res.* **2007**, *6*, 2351-2366.
- (12) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins; *J. Mol. Biol.* **2004**, *338*, 1015-1026.
- (13) Neduva, V.; Russell, R. B. Linear motifs: Evolutionary interaction switches; *Febs Lett.* **2005**, *579*, 3342-3345.
- (14) Hsu, W. L.; Oldfield, C. J.; Xue, B.; Meng, J.; Huang, F.; Romero, P.; Uversky, V. N.; Dunker, A. K. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding; *Prot. Sci.* **2013**, *22*, 258-273.
- (15) Jensen, M. R.; Ruigrok, R. W. H.; Blackledge, M. Describing intrinsically disordered proteins at atomistic resolution by NMR; *Curr. Opinion Struct. Biol.* **2013**, <http://dx.doi.org/10.1016/j.sbi.2013.1002.1007>.
- (16) Platzer, G.; Schedbauer, A.; Chemeli, A.; Ozdowy, P.; Coudeville, N.; Auer, R.; Kontaxis, G.; Hartl, M.; Miles, A. J.; Wallace, B. A.; Glaser, O.; Bister, K.; Konrat, R. The Metastasis-Associated Extracellular Matrix Protein Osteopontin Forms Transient Structure in Ligand Interaction Sites; *Biochemistry* **2011**, *50*, 6113-6124.
- (17) Holde, K. E. v.; Johnson, W. C.; Ho, P. S. *Principles of Physical Biochemistry*; Prentice Hall: New Jersey, 1998.
- (18) Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics; *Prot. Sci.* **2002**, *11*, 739-756.
- (19) Rao, J. N.; Jao, C. C.; Hegde, B. G.; Langen, R.; Ulmer, T. S. A Combinatorial NMR and EPR Approach for Evaluating the Structural Ensemble of Partially Folded Proteins; *J. Am. Chem. Soc.* **2010**, *132*, 8657-8668.
- (20) Schneider, R.; Huang, J.-r.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M. Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy; *Mol. BioSyst.* **2011**, DOI: 10.1039/c1mb05291h.
- (21) Drescher, M. EPR in Protein Science; *Top. Curr. Chem.* **2012**, *321*, 91-120.
- (22) Jeschke, G.; Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance; *Phys. Chem. Chem. Phys.* **2007**, *9*, 1895-1910.

- (23) Clore, G. M.; Tang, C.; Iwahara, J. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement; *Curr. Opinion Struct. Biol.* **2007**, *17*, 603-616.
- (24) Morin, B.; Bourhis, J. M.; Belle, V.; Woudstra, M.; Carriere, F.; Guigliarelli, B.; Fournel, A.; Longhi, S. Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling electron paramagnetic resonance spectroscopy; *J. Phys. Chem. B* **2006**, *110*, 20596-20608.
- (25) Schweiger, A.; Jeschke, G. *Principles of pulse electron paramagnetic resonance*; Oxford University Press: New York, 2001.
- (26) Jeschke, G.; Panek, G.; Godt, A.; Bender, A. Data Analysis Procedures for Pulse ELDOR Measurements of Broad Distance Distributions; Paulsen, H.; *Appl. Magn. Reson.* **2004**, *26*, 223-244.
- (27) Tanford, C. Isothermal unfolding of globular proteins in aqueous urea solutions; *J. Am. Chem. Soc.* **1964**, *86*, 2050-2059.
- (28) Sinha, N.; Smith-Gill, S. J. Electrostatics in Protein Binding and Function; *Curr. Prot. Peptide Sci.* **2002**, *3*, 601-614.
- (29) Avbelj, F.; Fele, L. Role of Main-chain Electrostatics, Hydrophobic Effect and Side-chain Conformational Entropy in Determining the Secondary Structure of Proteins; *J. Mol. Biol.* **1998**, *279*, 665-684.
- (30) Candotti, M.; Esteban-Martin, S.; Salvatella, X.; Orozco, M. Toward an atomistic description of the urea-denatured state of proteins; *Proc. Nat. Acad. Sci.* **2013**, *110*, 5933-5938.
- (31) Yang, Z. X.; Xiu, P.; Shi, B. Y.; Hua, L.; Zhou, R. H. Coherent Microscopic Picture for Urea-Induced Denaturation of Proteins; *J. Phys. Chem. B* **2012**, *116*, 8856-8862.
- (32) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: New York, 2003.
- (33) Huang, J. R.; Gabel, F.; Jensen, M. R.; Grzesiek, S.; Blackledge, M. Sequence-Specific Mapping of the Interaction between Urea and Unfolded Ubiquitin from Ensemble Analysis of NMR and Small Angle Scattering Data; *J. Am. Chem. Soc.* **2012**, *134*, 4429-4436.
- (34) Yi, Q.; Scalley-Kim, M. L.; Alm, E. J.; Baker, D. NMR characterization of residual structure in the denatured state of protein L; *J. Mol. Biol.* **2000**, *299*, 1341-1351.
- (35) Mittag, T.; Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles; *Curr. Opinion Struct. Biol.* **2007**, *17*, 3-14.
- (36) Uversky, V. N.; Santambrogio, C.; Brocca, S.; Grandori, R. Length-dependent compaction of intrinsically disordered proteins; *Febs Lett.* **2012**, *586*, 70-73.

- (37) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm; *J. Mol. Biol.* **1999**, *293*, 321-331.
- (38) Tompa, P.; Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions; *Trends Biochem. Sci.* **2008**, *33*, 2-8.
- (39) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis; *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 98-104.
- (40) Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition; *Nat. Chem. Biol.* **2009**, *5*, 789-796.
- (41) Grunberg, R.; Leckner, J.; Nilges, M. Complementarity of structure ensembles in protein-protein binding; *Structure* **2004**, *12*, 2125-2136.
- (42) Tompa, P.; Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions; *Trends Biochem. Sci.* **2008**, *33*, 2-8.
- (43) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy; *Proc. Natl. Acad. Sci.* **2012**, *109*, 16155-16160.
- (44) Junk, M. J. N., Spiess, H. W., and Hinderberger, D. The Distribution of Fatty Acids Reveals the Functional Structure of Human Serum Albumin. *Angew. Chem., Int. Ed.* **2010**, *49*, 8755-8759
- (45) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. and Kay, L. E. Backbone dynamics of a free and phosphopeptidecomplexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **1994**, *33*, 5984-6003
- (46) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277-293
- (47) Means, G. E., and Feeney, R. E. Reductive alkylation of amino groups in proteins. *Biochemistry* **1968**, *7*, 2192-2201

Supporting Information

Contents:

1. Supplementary DEER Spectroscopy
2. Details on time domain calculations
3. Measurements on OPN Single Mutants

1. Supplementary DEER Spectroscopy

In Figure S1c Δ_{eff} is plotted for time traces calculated from $P(R)$ s of random coils of different length as the function of the root-mean-square end-to-end distance, $\langle R^2 \rangle^{1/2}$.¹ This essentially quantifies how an elongation of an interspin distance distribution translates into Δ_{eff} for very broad distance distributions. $P(R)$ for random coils is a Gaussian distance distribution and can be calculated as stated in eq. S1-S3. Note that this is true not only for random coils, but for every distribution that is symmetric around $\langle R^2 \rangle^{1/2}$. The functions of Δ_{eff} were calculated for our specific experimental setup with $\lambda = 0.516$. From Figure S1c one can deduce that Δ_{eff} decays exponentially with $\langle R^2 \rangle^{1/2}$ of random coils if Δ_{eff} is always sufficiently smaller than λ . This is true in our experiments, since $\Delta_{\text{eff}} < 0.45$ in all cases.

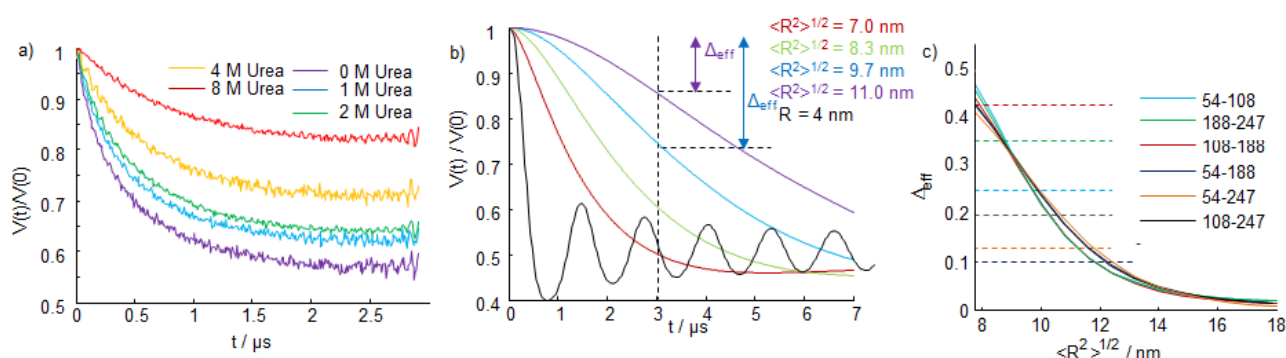


Figure S1. a) DEER time traces of C108-C188 at different urea concentrations. b) Calculated DEER time traces for different interspin distributions distances based on random coils with different $\langle R^2 \rangle^{1/2}$. Δ_{eff} is defined as the signal decay at $t = 3 \mu\text{s}$, as indicated by the double-headed arrows. The black time trace corresponds to a single interspin distance of 4 nm. c) Δ_{eff} as a function of $\langle R^2 \rangle^{1/2}$ of hypothetical random coil polypeptides (distance distribution) with segment numbers corresponding to the

residues between the labels of the different double mutants, n. n = 193 (54-247), 80 (108-188), 134 (45-188), 59 (188-247), 54 (54-108) and 139 (108-247). This maps how a certain mean distance translates into a Δ_{eff} -value but has no relevance for real random coils. The dashed lines indicate the maximum observed Δ_{eff} for a given double mutant. Corresponding time traces are similar to those shown in a). Note that the shift of $\langle R^2 \rangle^{1/2}$ at a given segment number was calculated as changes in Flory's characteristic ratio, c_{∞} .

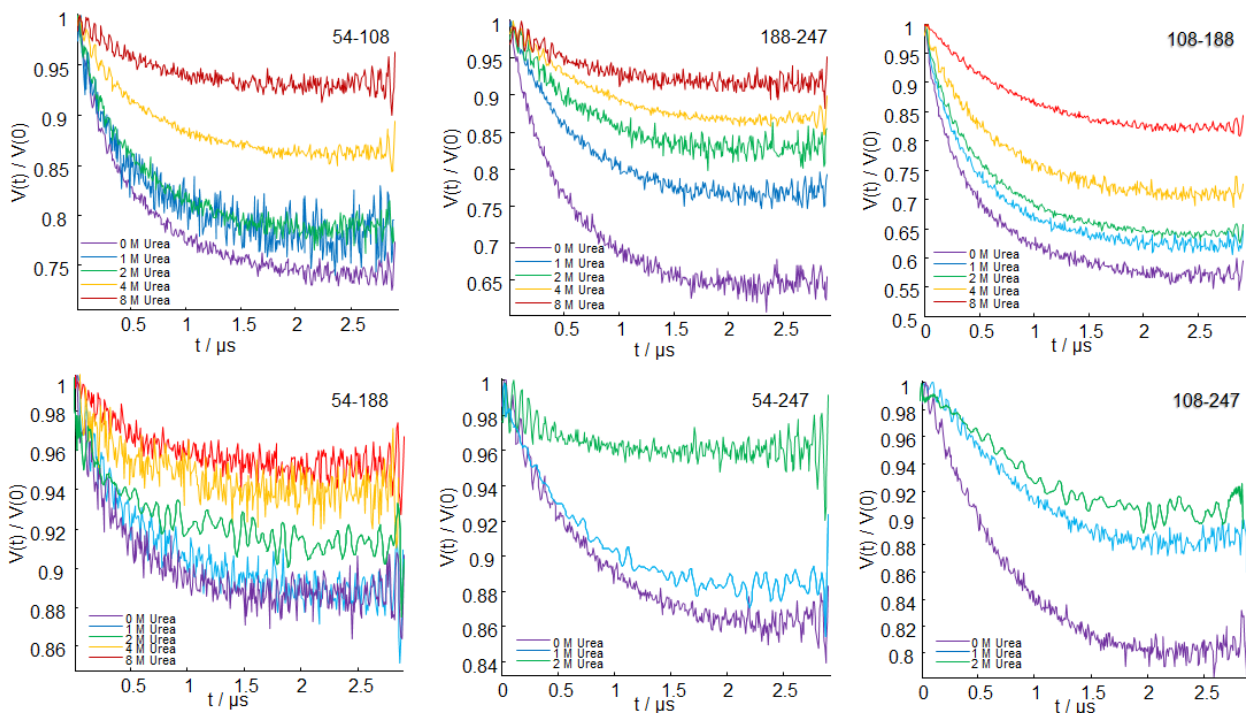


Figure S2. DEER time domain data for all double mutants for different urea concentrations. Absence of time traces indicates that we observed $\Delta_{\text{eff}} = 0$. Note the different scales of the $V(t)/V(0)$ axes.

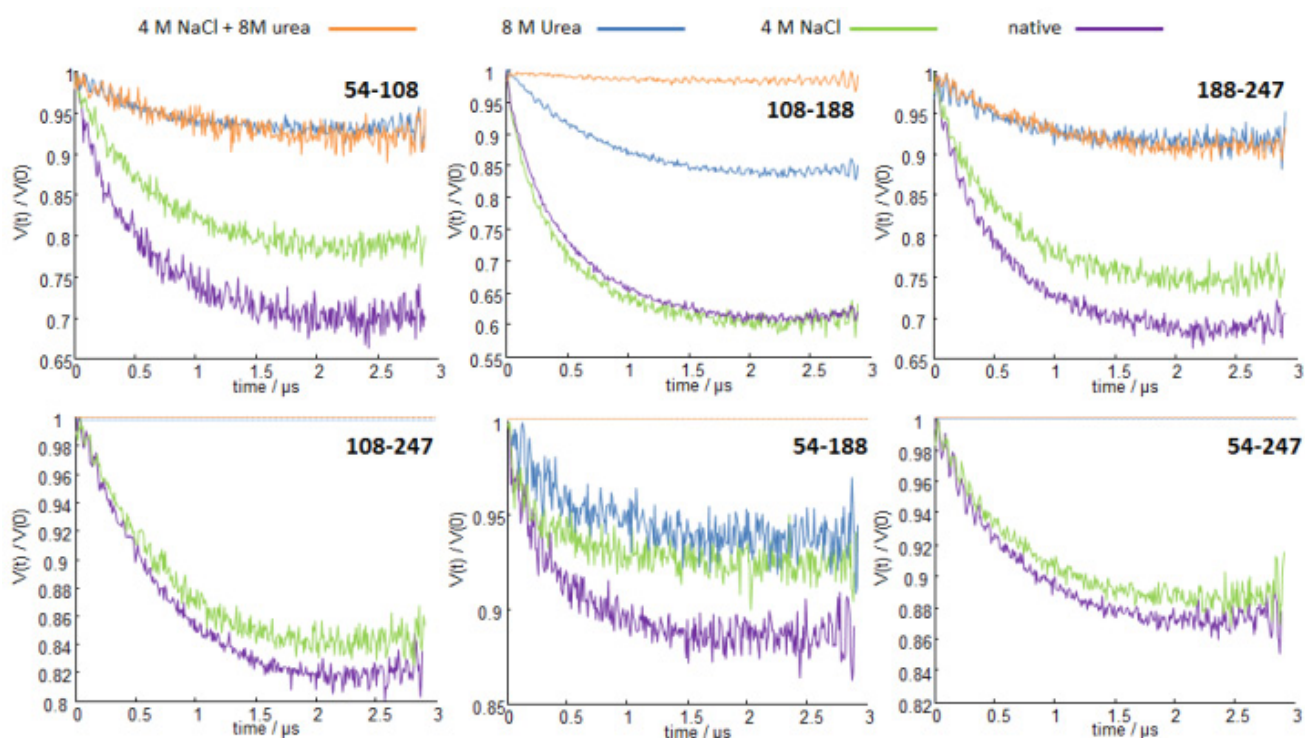


Figure S3. DEER time domain data for all double mutants for different denaturation conditions. Absence of time traces indicates that we observed $\Delta_{\text{eff}} = 0$. Note the different scales of the $V(t)/V(0)$ axes.

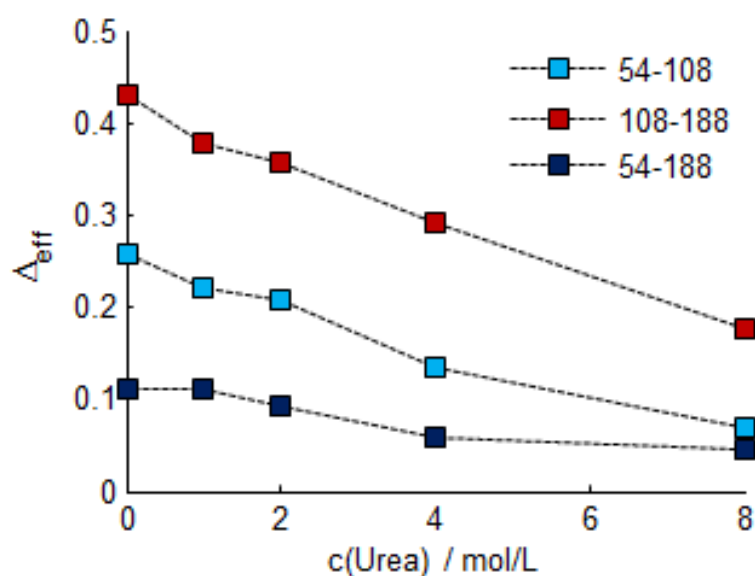


Figure S4. Δ_{eff} for several double mutants as a function of urea concentration.

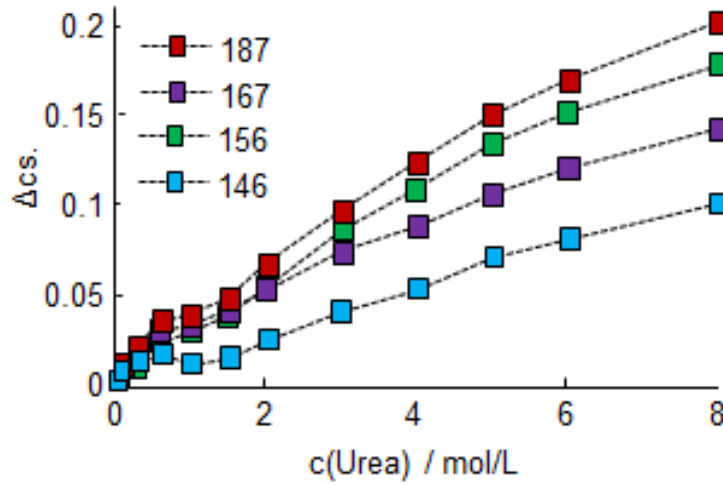


Figure S5. NMR chemical shift changes of selected residues as a function of urea concentration. Note that residue 146 shows a rather different behavior than the other three residues. However, like residue 146 also residues 156 and 167 are located in the central, more compact region of OPN.¹ Only 187 is located in a more flexible section of the IDP.

2. Details on time domain calculations (Figure S1b and c): All calculations were based on random coil models with:¹

$$\langle R^2 \rangle = c_{\infty} n l^2 \quad (1)$$

and

$$P(R, n, l) = \left(\frac{3}{2\pi n l^2} \right)^{3/2} \exp \left(-\frac{3(r - R)^2}{n l^2} \right) \quad (2)$$

Where $P(R, n, l)$ denotes the distribution of end-to-end vectors (see Figure S6a). n is the number of segments and l the segment length. The latter we assumed as the length of one amino acid, which is approximately 0.3 nm. DEER time traces were calculated as:

$$V(t) = \sum_R P(\omega_R) V_R(t) \quad (3)$$

with

$$V_R(t) = 1 - \int_0^{\pi/2} \lambda (1 - \cos(\omega_R t)) \sin \theta \, d\theta \quad (4)$$

θ is the angle between the external magnetic field and the vector connecting observer and pump spins. $P(\omega_R)$ is a distribution of dipolar coupling frequencies, ω_R , that corresponds to $P(R, n, l)$ as:

$$\omega_R = \frac{\mu_0 g_1 g_2 \mu_B}{4\pi \hbar R^3} (3 \cos^2 \theta - 1) \quad (5)$$

g_1 and g_2 denote the g-values of observer and pump spins. These were assumed to be equal. All other constants have their usual meanings. The contribution of compact conformations to $V(3 \mu s) = 1 - \Delta_{\text{eff}} V(3 \mu s) = 1 - \Delta_{\text{eff}}$ in dependence of $\langle R^2 \rangle$ can be estimated when calculating:

$$1 - \sum_{R_{\min}}^{R_{\max}} P(\omega_R) V_R(t) \bigg/ 1 - \sum_{R_{\min}}^{R(x)} P(\omega_R) V_R(t) \quad (6)$$

Where x denotes a hypothetic fraction of compact conformations of the overall conformational ensemble. Therefore, $R(x)$ has to fulfill $x = \int_{R_{\min}}^{R(x)} P(R, n, l) dr / \int_{R_{\min}}^{R_{\max}} P(R, n, l) dr$ (see Figure S6a). $R_{\min}$ is depending on the pump pulse length, τ_p , of a DEER experiment. For $\tau_p = 12$ ns $R_{\min} = 1.6 \text{ nm}$.² $R_{\max}$ is approximately 40 nm as calculated by Jeschke and co-workers.³ For $x = 0.05$, that is the cooperatively folded states are estimated to the most compact 5 % of $P(R, n, l)$ and for $t = 3 \mu s$ eq. S6 yields the function plotted in Figure S6b.

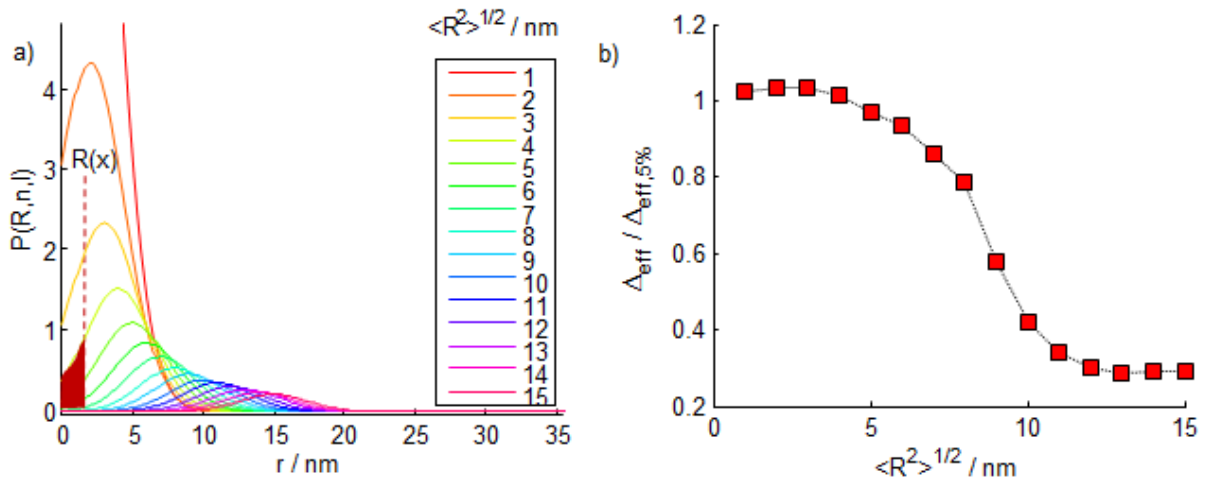


Figure S6. a) End-to-end vector distributions of random coils for different $\langle R^2 \rangle$. $R(x)$ is defined as the distance at which the integral between $R_{\min}$ and $R(x)$ matches $x\%$ of the total integral of $P(R, n, l)$. c_∞ was assumed as 2. b) Inverse contribution of the most compact 5% of conformations of the distributions in a) to Δ_{eff} . A value of 0 means that

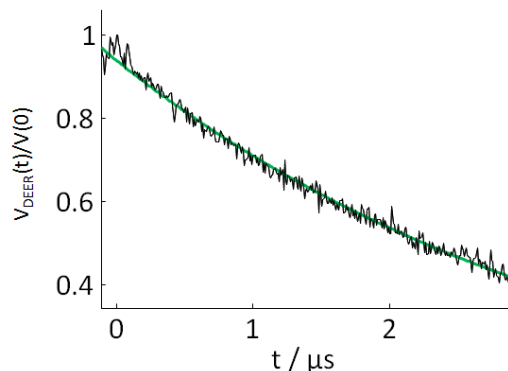
the DEER signal is totally dominated by the most compact 5 % of the conformational ensemble. Values below approximately $\langle R^2 \rangle^{1/2} = 5$ nm cannot be interpreted reliably, since most distances are below 1.6 nm, which is the lower sensitivity limit of the performed DEER measurements. Very long distances are subject to uncertainties, too, since the decay of the DEER signal becomes very shallow. Note that negative distances were treated by their absolute values in $P(R)$.

Judging from Figure S6, one can state that with increasing $\langle R^2 \rangle$ the contribution of compact conformations to Δ_{eff} increases. Yet, we only use this as a qualitative argument, since all the calculations are performed for random coil models and the most compact 5% of conformations are assumed as the cooperatively folded fraction. For actual OPN however $P(\omega_R)$ does not correspond to a random coil and x remains undetermined.

3. Measurements on OPN Single Mutants

To rule out dimerization of OPN and to gain information on the DEER background functions we performed DEER on the four single mutants C54, C108, C188 and C247. In Figure S7 uncorrected DEER data for C188 at 0.8 mM is shown. These data is representative also for C54, C108 and C247 and for combinations of all four mutants. In all cases we observed exponential decay functions that can be fitted with eq. 1 that is based on a homogeneous 3D distributions of spins as depicted in Figure S7.

Figure S7. Normalized, raw DEER time domain data for a 0.8 mM solution of single mutant C188 (representative also for C54, C108 and C247; these single mutant data were used for experimental background correction). The green fit corresponds to a homogeneous ($d = 3$) distribution of spins in the freeze-quenched solution (exponential decay). Hence, dimerization of OPN can be ruled out. The green line represents a fit based on a homogeneous exponential decay function.



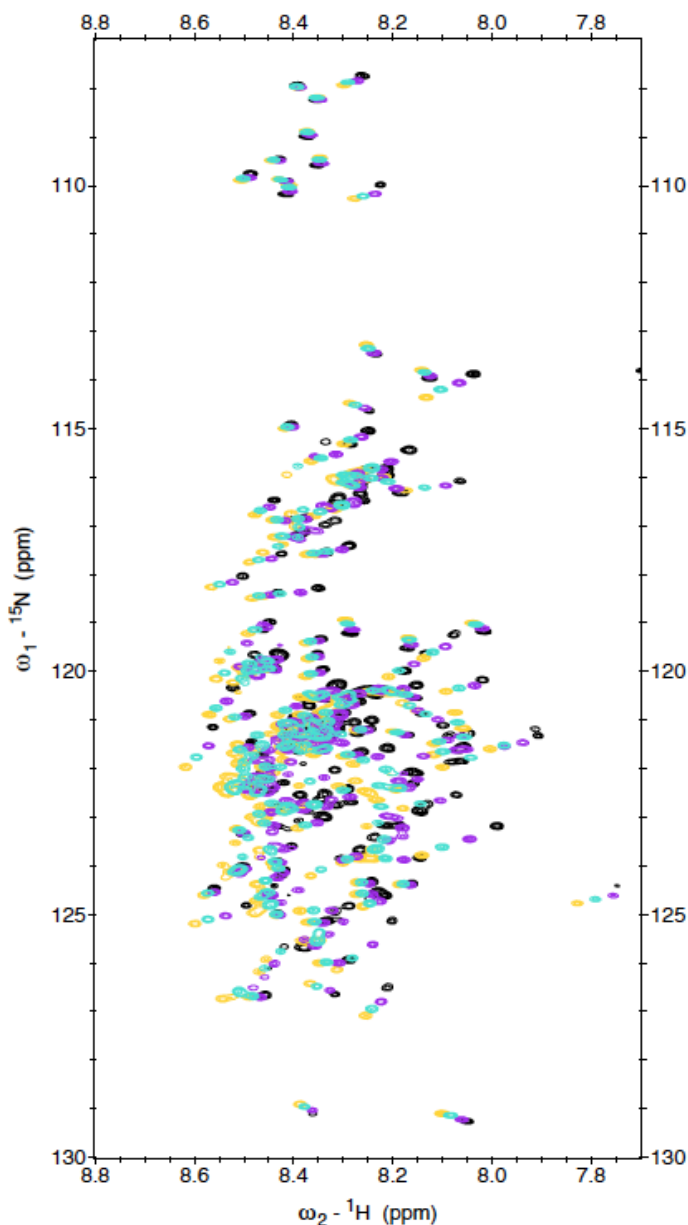


Figure S8. Full ^{15}N - ^1H HSQC spectra at different urea concentrations. Green: 0 M urea; pink: 2 M urea; blue: 4 M urea; black: 6 M urea

References

- (1) Rubinstein, M.; Colby, R. *Polymer Physics*; Oxford University Press: New York, 2004.
- (2) Tsvetkov, Y. D.; Milov, A. D.; Maryasov, A. G. Pulsed electron-electron double resonance (PELDOR) as EPR spectroscopy in nanometre range; *Russ. Chem. Rev.* **2008**, 77, 487-520.
- (3) Jeschke, G.; Koch, A.; Jonas, U.; Godt, A. Direct conversion of EPR dipolar time evolution data to distance distributions; *J Magn Reson* **2002**, 155, 72-82.

Chapter 2

Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex

Kurzbach, D.,* Schwarz, T.C.,* Platzer, G., Hoefler, S., Hinderberger, D., Konrat, R.
Angewandte Chemie International Edition **53**, 3840-3843 (2014)

*both authors contributed equally

ABSTRACT

Intrinsically disordered proteins (IDPs) play crucial roles in protein interaction networks and in this context frequently constitute important hubs and interfaces. Here we show by a combination of NMR and EPR spectroscopy that the binding of the cytokine osteopontin (OPN) to its natural ligand, heparin, is accompanied by thermodynamically compensating structural adaptations. The core segment of OPN expands upon binding. This “unfolding-upon-binding” is governed primarily through electrostatic interactions between heparin and charged patches along the protein backbone and compensates for entropic penalties due to heparin–OPN binding. It is shown how structural unfolding compensates for entropic losses through ligand binding in IDPs and elucidates the interplay between structure and thermodynamics of rapid substrate-binding and -release events in IDP interaction networks.

Results and Discussion

Intrinsically disordered proteins (IDPs) fulfill essential tasks in eukaryotic life despite a lack of well-defined structure. In particular, many IDPs are associated with versatile functions in protein interaction networks.^[1] It is their structural flexibility that allows them to adapt to and to interact with different binding partners, making IDPs suited for functioning as hubs between several interaction partners.^[2] However, this dynamic nature makes IDPs difficult to analyze. Nuclear magnetic resonance (NMR) has matured to a powerful experimental method for the characterization of IDPs and their complexes in solution.^[3] The binding modes between IDPs and their ligands nonetheless remain vaguely described. Different mechanisms have been distinguished: among them so-called “folding-upon-binding”,^[4] where the IDP adopts a certain conformation when interacting with a ligand, or the formation of fuzzy-complexes, where the IDP samples many conformations on the surface of a binding partner.^[5] Since IDPs are typically charged higher than globular proteins, electrostatics play a pronounced role in mediating the protein-ligand interactions of an IDP.^[6] Specific charge patterns of IDPs have been found to be optimized for attraction to and interaction with their naturally charged substrates.^[7] Here we provide by a combination

of NMR and electron paramagnetic resonance (EPR) data unprecedented insight into the subtleties of conformational adaptations occurring in IDP-substrate recognition events. While NMR measurements provide residue resolved information about structure (PRE; paramagnetic relaxation enhancement) and dynamics (NMR spin relaxation) EPR yields coarse-grained information about large-scale structural adaptations of IDPs.^[8] We analyze the interaction of the IDP Osteopontin (OPN), an extracellular matrix protein associated with metastasis of several kinds of cancer,^[9] with heparin,^[8f] a highly sulfated glycosaminoglycan widely used as anticoagulant. In a biological context heparin binding to OPN is of interest since it models the OPN-heparan sulfate and/or hyaluronic acid interaction, which is assumed to be involved in OPN-CD44 receptor (a heparan sulfate proteoglycan (HSPG) that sequesters various heparin-binding growth factors and chemokines) association - a process involved in cell signalling and adhesion.^[9b] Glycosaminoglycans like heparin, hyaluronan and heparan-sulfate are crucial for chemokine function *in vivo*.^[9c,d] Additionally, interactions between IDPs and biological polyelectrolytes are quite common^[1b] and our results might well be applicable to other systems. We show that upon binding to heparin OPN largely remains disordered although its structural ensemble is updated. The compensatory adaptations observed for OPN as a consequence of the heparin binding are mediated predominantly through electrostatic interactions and intriguingly demonstrate how entropic penalties due to local rigidification of binding site residues upon ligand binding are compensated through partial unfolding of peptide segments remote to the primary binding site. These results are important in the sense that they help to interpret the so far not well understood interplay between structure and thermodynamics of rapid substrate-binding and -release events in IDP interaction networks.

We investigate a 220 amino acids long mutant of OPN (residues 45-264 of native OPN plus an N-terminal Met). In Figure 1a) PRE changes upon heparin binding, Δ PRE, for four spin-labeled Cys-mutants of the truncated IDP are shown (labeling with MTSL; S-(2,2,5,5-tetramethyl-2,5-dihydro-¹H-pyrrol-3-yl)methyl methane-sulfonothioate, mutants C54, C108, C188 and C247). Δ PREs are calculated through changes in the relative (labeled/wildtype) ¹⁵N-¹H HSQC signal intensity ratios (height) upon heparin binding of OPN (MTSL protein interactions were excluded by reference measurements; see the Supporting Information). When interpreting PRE values of

IDPs one should be aware that due to the rapid conformational sampling of IDPs one observes ensemble averaged PRE data.^[8a] Hence, all conclusions drawn from these refer to “average” conformations and $\Delta\text{PRE} > 0$ indicates “on average” increasing distance between labeling site and a residue upon binding, $\Delta\text{PRE} < 0$ the opposite.^[8f] As can be observed the IDP displays differential changes of long-range backbone interactions as heparin binds. Specifically, spin labels attached to residues C108 and C188 experience displacements from the core region, that is, from residues 120-200 for C108 and from 70-120 for C188. Concluding, the PRE data indicate that the central segment (residues 90-150) of OPN expands upon heparin binding, comparable to an “unfolding-upon-binding” event. Intermolecular PRE contributions have been ruled out by a control experiment using a mixture of ^{14}N -labelled MTSL-tagged and ^{15}N -labelled untagged OPN (see the Supporting Information).

Differential structural compactness changes of the central region of OPN are further reflected in its motional dynamics. These were monitored by ^{15}N relaxation measurements (^{15}N longitudinal relaxation rates R_1 and transverse relaxation rates R_2 and ^{15}N - ^1H heteronuclear Overhauser enhancement, NOE) both in the apo (ligand free) and heparin-bound state. Figure 1b shows a residue plot of ^{15}N R_2 and ^{15}N - ^1H heteronuclear NOE differences. The regions containing most of the heparin binding site (140-170 and 180-200) display a decrease in ^{15}N R_2 due to local rigidification of the backbone upon binding (motions on the nanosecond time scale). This is paralleled by larger ^{15}N - ^1H heteronuclear NOE values (Figure 1b bottom) indicating reduced fast backbone motions on the picosecond timescale. In contrast, the region encompassing residues 90-120 exhibits increased backbone flexibility as evidenced by both, decreased R_2 values and more negative heteronuclear NOEs. The finding that residues 90-120 are more flexible in the heparin-bound state is in good agreement with decreased PRE values and further validates that this part of the protein becomes more displaced from the heparin binding site and thus less restrained upon complex formation. Overall, the NMR relaxation data demonstrates that OPN displays more conformational flexibility in the bound state. The binding of heparin was further confirmed by DOSY (diffusion ordered spectroscopy) NMR that showed an increase of the hydrodynamic radius of OPN through heparin binding from 3.6 nm to 5.2 nm.

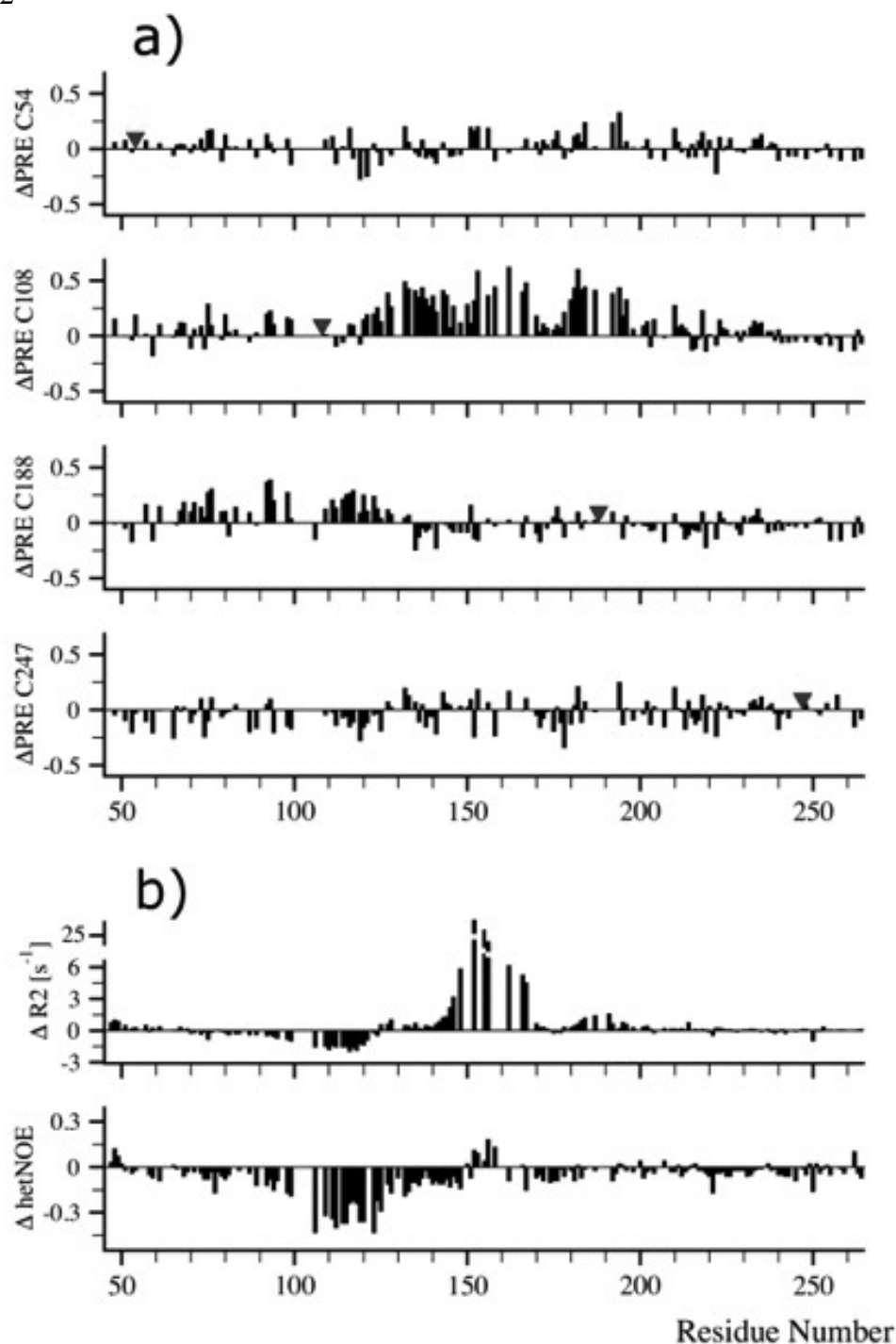


Figure 1. a) ΔPRE for the four different labeling sites upon heparin binding from ^{15}N - ^1H HSQC signal intensities. (ΔPRE = relative (OPN-MTSL/OPN) ^{15}N - ^1H HSQC-Intensity (heparin-bound) minus the relative ^{15}N - ^1H HSQC-Intensity (apo)). ΔPRE values > 0 indicate increasing distances to the labeling site (gray triangles mark the spin-labeled sites); ΔPRE values < 0 indicate decreasing distances. For error estimates see the Supporting Information. b) ΔR_2 and ΔhetNOE (^{15}N - ^1H NOE changes (bound minus apo)) as a function of residue position. For error estimates see Supporting Information.

To complement the NMR-derived data on OPN we performed both room temperature continuous wave (CW) EPR on the four available single mutants and double electron-electron resonance (DEER) measurements on six spin-labelled Cys double-mutants of the truncation mutant. (DEER was performed at 50 K on shock-frozen samples that contain the conformational protein ensemble present at the glass transition temperature of the solution, which is assumed to be representative for the solution structural ensemble. For further details see the Supporting Information.) CW EPR reveals that the label attached to residue C108 gains mobility through heparin binding, while the mutant C188 exhibits a more restricted rotational motion (with rotational correlations times of $\tau_{c,108}=0.52$ ns, $\tau_{c,108/\text{Heparin}}=0.43$ ns, $\tau_{c,188}=0.69$ ns and $\tau_{c,188/\text{Heparin}}=0.87$ ns; see the Supporting Information). τ_c values of spin labels attached to residues C54 and C247 appear not to be affected significantly by heparin-binding. Again, the increasing rotational mobility of MTSL at C108 is indicative for pronounced solvent exposure of this labeling site as a consequence of the expansion of the core segment of OPN.^[10] Note that τ_c s derived from CW EPR suffer from the nonseparability of local and global rotational correlation times (similar to NMR data) and only represent an effective correlation time corresponding to the nitroxide moiety motion. To unequivocally probe changes of rotational correlation times due to binding of the ligand both NMR (R_2 and HetNOE) and EPR data are required.

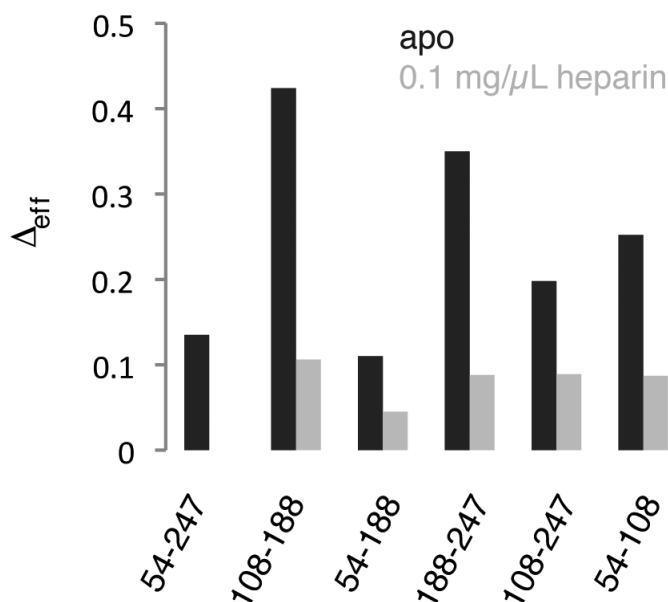


Figure 2. Δ_{eff} values for the six OPN double mutants in the apo (black) and heparin-bound (gray) state. (Δ_{eff} for mutant 54-247 with heparin is 0.)

While detailed characterization and analysis of the DEER time traces of the OPN apo state have already been reported,^[11] here the DEER data were recorded in the absence and the presence of heparin. All DEER data are shown in the Supporting Information. As described before, DEER data of IDPs with broad conformational ensembles are difficult to analyze due to a lack of signal modulation.^[8d, 11] Thus, since the standard DEER analysis techniques fail for OPN, time traces were analyzed using an effective modulation depth, Δ_{eff} , as introduced in our previous work.^[11] Δ_{eff} denotes the total signal decay at the experimental DEER evolution time (see the Supporting Information) and is an approximate measure of the average interspin distance for broad $P(R)$ distributions.^[11] For increasing interspin mean distance R between two labeling sites Δ_{eff} typically decreases. The lower distance boundary of this method is 1-1.6 nm in aqueous solutions and the assumption of decreasing Δ_{eff} with increasing mean distance of $P(R)$ is only valid for double mutants showing very broad conformational ensembles and, hence, very broad distance distributions, $P(R)$.^[11] Thus, our approach is appropriate primarily for the analysis of large, extended proteins like OPN. (Note that PREs are complementary to DEER in terms of length scales. PREs are effective below 2-3nm only for the present system. Thus, also significant changes in Δ_{eff} do not need to correlate with changes in PREs or differential PRE values.) Furthermore, experimental DEER background functions are necessary for proper data treatment. It should be noted that for folded proteins where one frequently observes strong modulations in DEER time traces a Δ_{eff} analysis would not be possible. In Figure 2 Δ_{eff} values for the six double mutants in the apo and heparin-bound state are shown (the EPR raw data are shown in the Supporting Information Figures S2 and S3). In all cases binding of heparin leads to a clear decrease in Δ_{eff} indicating longer distances between the labeling sites of the six double mutants. This is in excellent agreement with the expansion of OPN upon heparin binding, observed by means of NMR analysis and also supports an unfolding-upon-binding event.

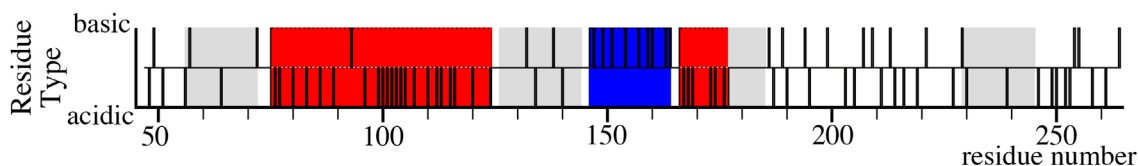


Figure 3. Charge map of OPN. Grey indicates charge neutral patches, red indicates acidic (negatively charged under experimental pH) residues and blue basic (positively charged) residues.

The observed, “on average” conformational expansion of the central region of OPN upon heparin binding can be rationalized when taking into account that heparin is a highly, negatively charged polyelectrolyte that interacts predominantly electrostatically with its environment. In Figure 3a the charge distribution map of OPN is shown which highlights patches of low and high charge densities. Heparin has been shown to bind to the positively charged patch of OPN around residues 155, thereby exploiting favourable electrostatic interactions.^[8f, 11] Yet, the large negatively charged patch around residues 80-120 is likely to experience electrostatic repulsion from heparin and is, therefore, displaced from the central region of OPN around residue 155 upon heparin binding. This leads to the observed expansion of the central region and to positive Δ PREs as shown in Figure 1a. In contrast, in the apo state the negatively charged region around C108 is known to be attracted by the positive patch around residue 155, compacting the OPN core.^[11] Thus, the OPN-heparin complex reveals an electrostatic binding mode largely governed by complementary charge patterns as often found in molecular recognition processes.^[12] To further corroborate these findings we measured the changes in NMR chemical shifts, PRE data (see Figure S6 in the Supporting Information) and DOSY data caused by different salt conditions. The observed changes of NMR parameters and hydrodynamic radius R_H under high salt conditions clearly indicate that the OPN-heparin complex dissociates through NaCl influence (R_H OPN: 3.6 nm; OPN-Heparin: 5.2 nm and OPN-Heparin/400 mM NaCl: 3.4 nm).

The binding of heparin to OPN was also studied by isothermal titration calorimetry (ITC) measurements. From ITC measurements ΔH and ΔS values of -16.3 kCal mol⁻¹ and -35 Cal mol⁻¹ K⁻¹ can be obtained for the heparin-OPN binding event assuming an average molecular weight of 17.5 kDa for heparin. These quite large ΔH

and ΔS values nearly cancel each other at 293 K with respect to the Gibbs energy of binding ($\Delta G = \Delta H - T\Delta S$). The dissociation constant was determined to 34 μM (see the Supporting Information). Additionally, ITC measurements at different temperatures revealed a positive differential heat capacity, ΔC_p , (about 40 $\text{cal M}^{-1}\text{K}^{-1}$) for heparin binding which further supports the notion that the bound state exhibits more conformational flexibility compared to the free state. Again, ITC data obtained at 400 mM NaCl confirmed the inhibition of the complex formation under high salt conditions (see Figure S7 in the Supporting Information).

In conclusion we could show, by applying NMR and EPR measurements, that heparin binding leads to an expansion of the central regions of OPN. Therefore, the interaction mode of this IDP with its ligand can be described as formation of a fuzzy complex (i.e., residual conformational flexibility in the bound state). Although there is local rigidification in the heparin binding cleft (region around residue 155) the resulting conformational entropy penalty is reduced by a compensatory increase in conformational flexibility of the negatively charged region 90-120. The local (or segmental) unfolding and expansion of the OPN core segment thereby significantly contributes to the overall thermodynamic equilibrium balanced between counteracting contributions like solvation enthalpy and rotational and translational degrees of freedom and conformational entropy, as typical for IDPs.^[6, 14]

We deduce that heparin binding to OPN is predominantly mediated via electrostatic interactions across the interface displaying complementary charge distributions. Given its similar chemical composition an analogous binding mode can be anticipated for the naturally occurring OPN ligand heparan sulfate. Further, the thermodynamic (entropic) compensation, as measured by NMR and ITC, gained through the structural rearrangements of OPN might be central to the common IDP property of rapid substrate binding and release and the versatility of their host-guest interactions. It is thus instructive to compare our findings with a recent study describing the binding mode of Sic1 to Cdc4 for which a similar binding mode was observed.^[15] In this study conformational averaging of Sic1 in the bound state was shown to be relevant for polyvalent interactions that lead to ultrasensitivity and nonlinear binding in response to, for example, phosphorylation. Although the details of binding differ (OPN-heparin: bulk electrostatic patches versus Sic1-Cdc4: polyvalent

interactions) entropic compensatory events seem to be present in both systems (OPN-heparin: conformational expansion; Sic1-Cdc4: enhanced conformational entropy through polyvalent interaction) hinting towards a more common phenomenon in the realm of IDPs.

IDPs mediate protein interactions in dynamic networks; for fast and efficient response to external stimuli low energy barriers and facilitated interconversions between different substates are required. This is facilitated by binding modes governed by electrostatics since they allow for considerable conformational plasticity (also through high hydrophilicity),^[16] modulating the lifetime and rates of conversions of individual bound states in encounter complex ensembles. We thus anticipate that complementary structural and dynamical adaptations similar to the described OPN-heparin interaction will be observed for intrinsically disordered protein hubs in cellular interaction networks.

Acknowledgements, Funding

We thank Christian Bauer for technical support and Prof. Hans W. Spiess for continuing support. D.K. acknowledges a Feodor-Lynen Scholarship from the Humboldt Foundation and support by the Gutenberg Academy of the University of Mainz. This work was supported by FWF (W-1221-B03). The expression plasmid for wildtype OPN was provided by the group of Prof. K. Bister (University of Innsbruck).

References

- [1] a) P. Tompa, *Trends Biochem. Sci.* **2002**, 27, 527-533; b) J. Dyson, P. E. Wright, *Nat. Rev.* **2005**, 6, 197-208.
- [2] D. D. Boehr, R. Nussinov, P. E. Wright, *Nat. Chem. Biol.* **2009**, 5, 789-796.
- [3] a) P. R. L. Markwick, T. Malliavin, M. Nilges, *Plos Comput. Biol.* **2008**, 4, e100168; b) G. T. Montelione, D. Y. Zheng, Y. P. J. Huang, K. C. Gunsalus, T. Szyperski, *Nat. Struct. Biol.* **2000**, 7, 982-985; c) V. Ozenne, R. Schneider, M. Yao, J.-r. Huang, L. Salmon, M. Zweckstetter, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2012**, 134, 15138-15148.
- [4] H. J. Dyson, P. E. Wright, *Curr. Opinion Struct. Biol.* **2002**, 12, 54-60.

- [5] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2-8.
- [6] V. N. Uversky, *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1090-1103.
- [7] a) T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, J. D. Forman-kay, *Structure*, **2010**, *18*, 494-506; b) R. Konrat, *Structure*, **2010**, *18*, 416-149.
- [8] a) G. M. Clore, C. Tang, J. Iwahara, *Curr. Opinion Struct. Biol.* **2007**, *17*, 603-616; b) G. Jeschke, Y. Polyhach, *PCCP*, **2007**, *9*, 1895-1910; c) J. Iwahara, C. D. Schwieters, G. M. Clore, *J. Am. Chem. Soc.* **2004**, *126*, 5879-5896; d) M. Drescher, *Top. Curr. Chem.* **2012**, *321*, 91-120; e) T. Bund, J. M. Boggs, G. Harauz, N. Hellmann, D. Hinderberger, *Biophys. J.* **2010**, *99*, 3020-3028; f) G. Platzer, A. Schedbauer, A. Chemeli, P. Ozdowy, N. Coudeville, R. Auer, G. Kontaxis, M. Hartl, A. J. Miles, B. A. Wallace, O. Glaser, K. Bister, R. Konrat, *Biochemistry*, **2011**, *50*, 6113-6124.
- [9] A. Bellahcène, V. Castronovo, K. U. E. Ogbureke, L. W. Fisher, N. S. Fedarko, *Nat. Rev.* **2008**, *8*, 212-226. b) H. Ponta, L. Sherman, P. A. Herrlich, *Nat. Rev.* **2003**, *4*, 33-45. c) M. Jones, L. Tussey, N. Athanasou, D. G. Jackson, *J. Biol. Chem.* **2000**, *275*, 7964-7794. d) J. L. de Paz, E. A. Moseman, C. Noti, L. Polito, U. H. von Andrian, P. H. Seeberger, *ACS Chem. Biol.* **2007**, *2*, 735-744.
- [10] a) W. Hubbell, D. Cafiso, C. Altenbach, *Nat. Struct. Biol.* **2000**, *7*, 735-739; b) W. Hubbell, H. Mchaourab, C. Altenbach, M. Lietzow, *Structure*, **1996**, *4*, 779-783.
- [11] D. Kurzbach, G. Platzer, M. Henen, T. Schwarz, R. Konrat, D. Hinderberger, *Biochemistry*, **2013**, *52*, 5167-5175.
- [12] a) B. Honig, A. Nicholls, *Science* **1995**, *268*, 1144-1149; b) N. Sinha, S. J. Smith-Gill, *Curr. Prot. Peptide Sci.* **2002**, *3*, 601-614.
- [13] D. Hinderberger, H. W. Spiess, G. Jeschke, *Appl. Magn. Reson.* **2010**, *37*, 657-683.
- [14] K. K. Frederick, M. S. Marlow, K. G. Valentine, A. J. Wand, *Nature*, **2007**, *448*, 325-U323.
- [15] a) M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay, H. S. Chan, *Proc. Nat. Acad. Sci.* **2007**, *104*, 9650-9655. b) T. Mittag, S. Orlicky, W. -Y. Choy, X. Tang, H. Lin, F. Sicheri, L. E. Kay, M. Tyers, J. D. Forman-Kay, *Proc. Nat. Acad. Sci.* **2008**, *46*, 17772-17777.
- [16] D. Kurzbach, W. Hassounah, J. R. McDaniel, E. A. Jaumann, A. Chilkoti, D. Hinderberger, *J. Am. Chem. Soc.* **2013**, *135*, 11299-11308.

Supporting Information

Protein Preparation

The expression and purification of recombinant quail OPN protein (OPN220) were performed as described previously.^[1] Cysteine mutations were introduced using the QuickChangeII site directed mutagenesis kit (Stratagene). For NMR and EPR analysis, all protein samples were concentrated to 0.8 mM in phosphate buffer (50mM sodium phosphate, 50 mM NaCl, pH 6.5) in a 90% H₂O / D₂O mixture. Heparin was purchased from Sigma Aldrich (Product Nr. H4784) and was added to a final concentration of 0.1 mg/μl (EPR) or 15 mg/mL (NMR) where present.

DEER

DEER is applied to glassy solids obtained by freeze-quenching the solutions after addition of 30 v/v % glycerol. (Note that glycerol typically does not influence protein structure.^[2]) This is achieved by immersing the sample tube in supercooled isopentane. Such, a snapshot representative for the solution at the glass transition temperature is detected, as evidenced by manz previous DEER studies on proteins.^[2] The sample volume was always large enough to fill the complete resonator. The four pulse DEER sequence $\pi/2(\nu_{\text{obs}}) - \tau_1 - \pi(\nu_{\text{obs}}) - (\tau_1 + t) - \pi(\nu_{\text{pump}}) - (\tau_2 - t) - \pi(\nu_{\text{obs}}) - \tau_2$ - echo was used to obtain dipolar time evolution data at X-band frequencies (9.2 to 9.4 GHz) with a Bruker Eleksys 580 spectrometer equipped with a Bruker Flexline split-ring resonator ER4118X_MS3. The dipolar evolution time t was varied, whereas $\tau_2 = 2$ μs and τ_1 were kept constant. Proton modulation was averaged by the addition of eight time traces of variable τ_1 , starting with $\tau_{1,0} = 200$ ns and incrementing by $\Delta\tau_1 = 8$ ns. The resonator was overcoupled to $Q \approx 100$. The pump frequency, ν_{pump} , was set to the maximum of the EPR spectrum. The observer frequency, ν_{obs} , was set to $\nu_{\text{pump}} + 61.6$ MHz, coinciding with the lowfield local maximum of the nitroxide spectrum. The observer pulse lengths were 32 ns for both $\pi/2$ and π pulses, and the pump pulse length was 12 ns. The temperature was set to 50 K by cooling with a closed cycle cryostat (ARS AF204, customized for pulse EPR, ARS, Macungie, PA). The resulting time traces were normalized to $t = 0$. Background correction was done through dividing by

experimental functions gained from DEER on the four single mutants. These functions correspond to a homogeneous exponential decay and to a homogeneous ($d = 3$) distribution of spins.

NMR Spectroscopy

NMR spectra were recorded at 20°C on Varian spectrometers operating at 500, 600 and 800 MHz. OPN220 protein samples were dissolved to 0.8mM in 50mM sodium phosphate, 50mM NaCl at pH 6.5 with 10% D₂O as lock solvent. PRE intensity ratios were derived from PFG sensitivity enhanced two-dimensional ¹H-¹⁵N HSQC spectra of ¹⁵N labeled OPN220 mutants C54, C108, C188 and C247 respectively.^[3] NMR spectra were processed using NMRPipe^[4] and analyzed using SPARKY^[5]. DOSY measurements were performed in the presence of 0.5% dioxane at a protein concentration of 0.1 mM. Note that the determination of hydrodynamic radii may be influenced under high salt conditions by a reduced mobility of Dioxane.

Paramagnetic Relaxation Enhancements (PREs)

Single cysteine OPN220 mutants (C54, C108, C188 and C247) were tagged with the nitroxide spin-label (1-Oxyl-2,2,5,5-tetramethyl-Δ³-pyrroline-3-methyl) Methanethiosulfonate (MTSL). The overall PRE effect on OPN220 was measured as intensity ratio of cross peaks between the paramagnetically labeled mutants and electrochemically reduced counterparts. This was cross-checked by comparison with the intensity ratio of cross-peaks in presence (I_{+MTSL}) and absence (I_{-MTSL}) of the cysteine-attached spin labels, $\Delta_{MTSL} = I_{+MTSL}/I_{-MTSL}$. To account for possible interactions of the spin-label with the protein, MTSL was added to untagged, ¹⁵N labeled OPN220 at a final concentration of 1mM. The intensity ratios of OPN220 with unbound/free ($I_{+FREEMTSL}$) and without (I_{-MTSL}) MTSL, $\Delta_{UNSPEC} = I_{+FREEMTSL}/I_{-MTSL}$, were combined with the intensity ratios from attached MTSL (Δ_{MTSL}) to calculate the PRE effect on the protein: $PRE = \Delta_{MTSL} / (\Delta_{MTSL} + (1 - \Delta_{UNSPEC}))$. Further, intermolecular contributions could be ruled out by DEER on the single mutants, which showed a homogeneous distribution of the spins in OPN Cys-mutant solutions. Additionally, ¹⁴N OPN Mutants C54 and C247 were tagged with MTSL and each mixed with ¹⁵N OPN and Heparin (¹⁴N OPN 0.4 mM, ¹⁵N OPN 0.4 mM, Heparin 0.1 mg/ml) with no measurable PRE effect. Despite all these reference measurements it should be noted that every labeling study might always suffer from the drawback that the protein side

chains or backbone might interact with the label and thus bias the protein structure. However, HSQC measurements with a reduced (diamagnetic) label did not show any significant resonance shifts, when compared to the wild type. Thus, the labeling procedure does not seem to affect the conformational ensemble.

^{15}N $R_{1,2}$ and ^{15}N - ^1H NOE Spectra

Information on dynamics was obtained by determining backbone ^{15}N T_1 and ^{15}N T_2 relaxation times as described previously.^[3] All measurements were carried out at 20°C at a field corresponding to a ^1H Larmor Frequency of 599.89 MHz. Contributions of overlapping peaks were deconvoluted by peak fitting. To determine the ^{15}N transverse relaxation time T_2 Carr-Purcell-Meiboom-Gill (CPMG) delays of 0, 16, 32, 64, 96, 160, 256, 384 and 512 ms were employed using a CPMG duty cycle delay of 0.5ms. To determine the ^{15}N longitudinal relaxation time T_1 delays of 0, 20.8, 41.6, 83.2, 124.8, 208.0, 624, 1248 and 1996.8 ms were employed. Using an extension script of SPARKY^[5] the resulting series of cross-peak intensities was fitted to a two parameter exponential decay: $I(t) = A \exp[-(t/T)]$, where $I(t)$ represents the delay time-dependent peak intensity. Random error estimates for T_1 and T_2 were calculated with a best fit for a set of perturbed heights. $R_{1,2}$ was calculated as $1/T_{1,2}$. Perturbation of the heights was carried out with computer-simulated noise with a Gaussian distribution of zero mean original heights of the best fit and with variance equal to the root-mean-square deviation. Random experimental errors were determined as spread of the best fit values.

Heteronuclear steady-state NOE ^{15}N - $\{^1\text{H}^{\text{N}}\}$ attenuation factors were derived from the ratio of peak intensities of experiments with and without prior proton saturation $I_{\text{NOE}}/I_{\text{nonNOE}}$.^[3] Spectra determined under proton saturation were recorded with a 4 s relaxation delay prior to a 4.4 s proton presaturation period; non-saturated spectra featured a net relaxation delay of 8.4s. For all relaxation experiments acquisition times for the nitrogen and proton dimensions were 259ms and 96ms, respectively.

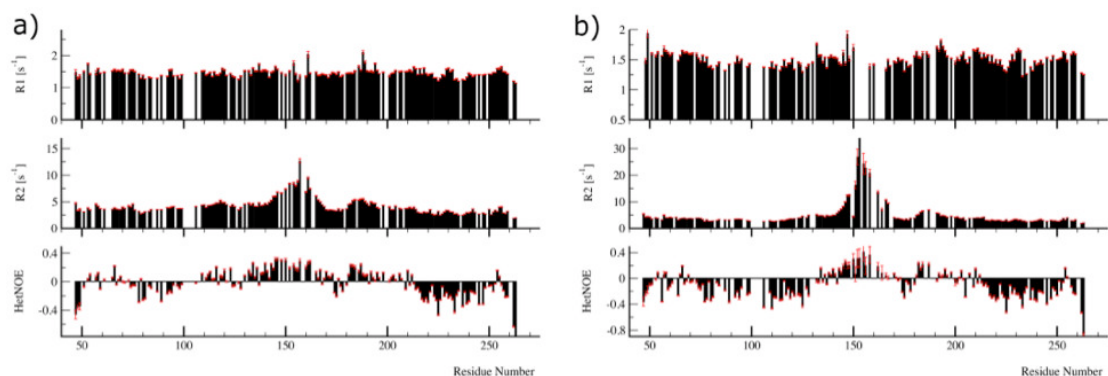


Figure S1. ^{15}N $R_{1,2}$ and ^{15}N - ^1H NOE values for OPN in its a) free and b) heparin-bound state. Please note the differences in regions 90-120 and 140-170 in both ^{15}N R_2 times and ^{15}N - ^1H NOE relative intensities (highlighted in Figure 1b) as ΔR_2 and ΔHetNOE). The red bars indicate experimental errors.

Supplementary DEER

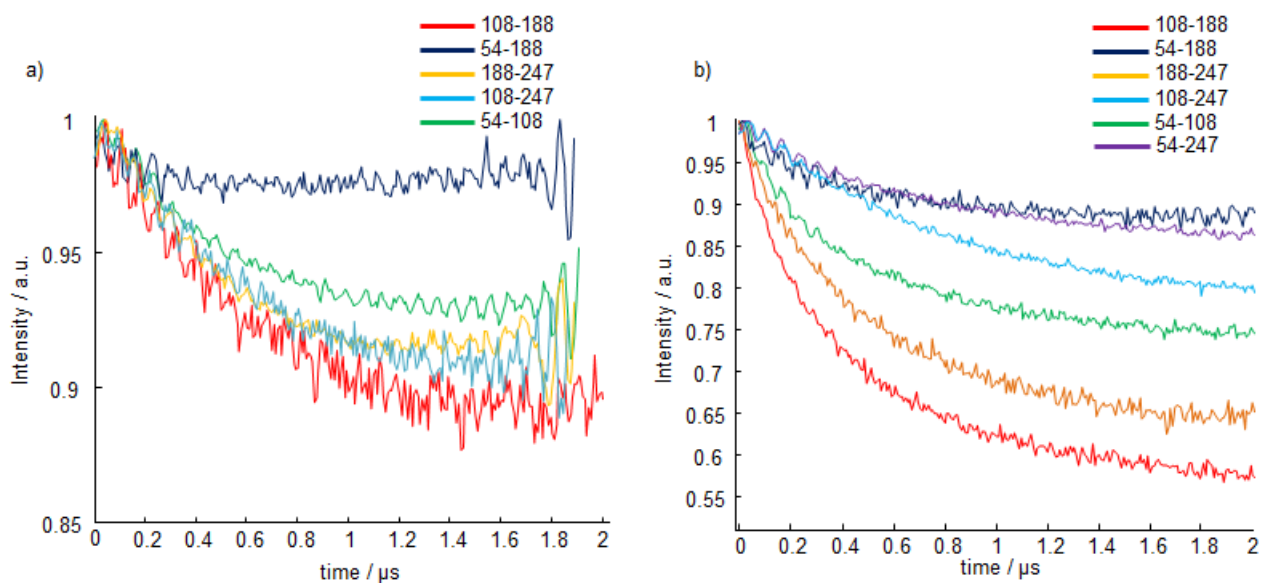


Figure S2. DEER time traces of the double mutants under investigation a) in the presence and b) in the absence of 0.1 mg/ μL heparin. Note the different intensity scales. Missing time traces denote $\Delta_{\text{eff}} = 0$.

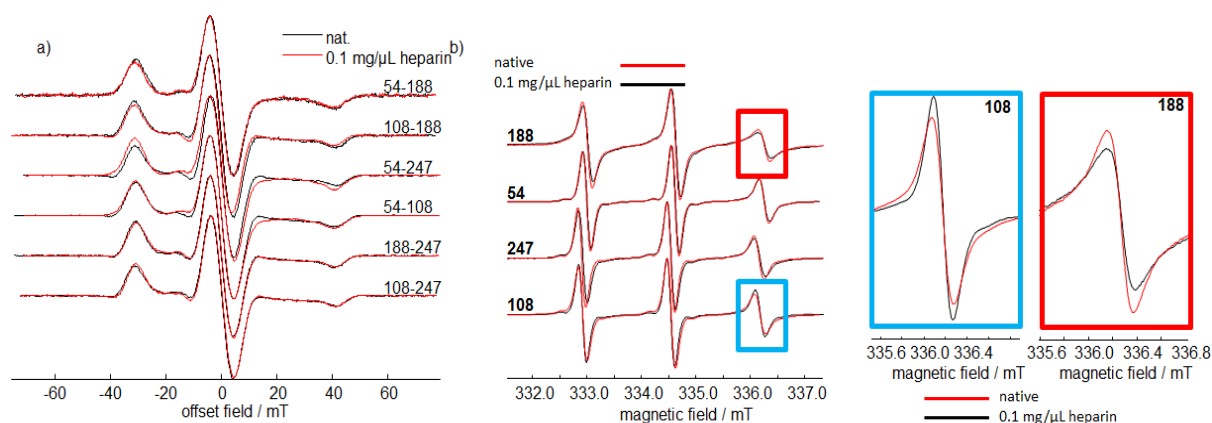


Figure S3. a) Pseudo modulated low temperature spin echoes of the double-mutants under investigation in the absence (black) and presence (red) of 0.1 mg/μL heparin. The modulation frequency was 3 mT. b) Room temperature CW EPR spectra of the four single-mutants under investigation in the absence (red) and presence (black) of 0.1 mg/μL heparin and a zoom on the high-field transitions of C108 (blue) and C188 (red).

Supplementary ITC

Preparation and Analysis of Solutions – Although the molecular weight of heparin can differ between 6 kDa and 30 kDa, commercially available heparin, such as the one used, have most chains in the range of 16-19 kDa (*Sigma supplier data, quality control information*) and are, therefore, amenable to dialysis. Heparin solutions were prepared (calculating with an average weight of 17.5kDa) in deionized Milli-Q water. Both Heparin and Osteopontin were dialyzed over night at 4°C into the same degassed buffer (50mM Na₂HPO₄, 50mM NaCl, 1mM Azide, 1mM EDTA, pH=6.5). The concentrations of OPN and Heparin were lowered to the desired value using the dialysis buffer and both solutions were then centrifuged at 36.000 g for 20 min. The correct concentration of OPN was confirmed by measuring OD₂₈₀.

Measurements and Analysis – Titration calorimetric measurements were carried out using a MicrocalTM iTC200 calorimeter. The product of the initial ligand concentration and the association constant $c = K_b \cdot [\text{Hep}]$ was in the range of $2 < c < 200$ as required for ITC studies.^[6] The starting conditions were 130-150 μM OPN in the measurement cell (200μL) and 1.5mM Heparin in the injection needle (40 μL). Titration of Heparin into buffer gave only minor heat release. Nevertheless these titrations were subtracted from runs performed with OPN. Measurements were carried

out at 283.2 K, 293.2 K, 303.2 K and 313.2 K with a stirring speed of 1000 rpm. The time delay between peaks was set to 180s to allow peak recovery to baseline. The injection scheme and concentrations were chosen so that a flat start and saturation could be obtained (Figure S4) and the areas below start and end peaks were reduced by more than a factor of 10 (corrected for different injection volumes). The injection scheme was set to $8 \cdot 0.5 \mu\text{L} + 8 \cdot 1 \mu\text{L} + 13 \cdot 2 \mu\text{L}$ (with the injections lasting 1, 2 and 4s respectively). The first $0.5 \mu\text{L}$ injection was not used in the fitting process. To perform a nonlinear least-squares fit of the experimental data, the ‘one set of sites’ model within the OriginTM Program^[7] was chosen. Averaging 3 measurements we obtained values of -16.3 kCal/mol, -35 Cal/mol/K and 29733 mol^{-1} for ΔH , ΔS and K_b respectively.

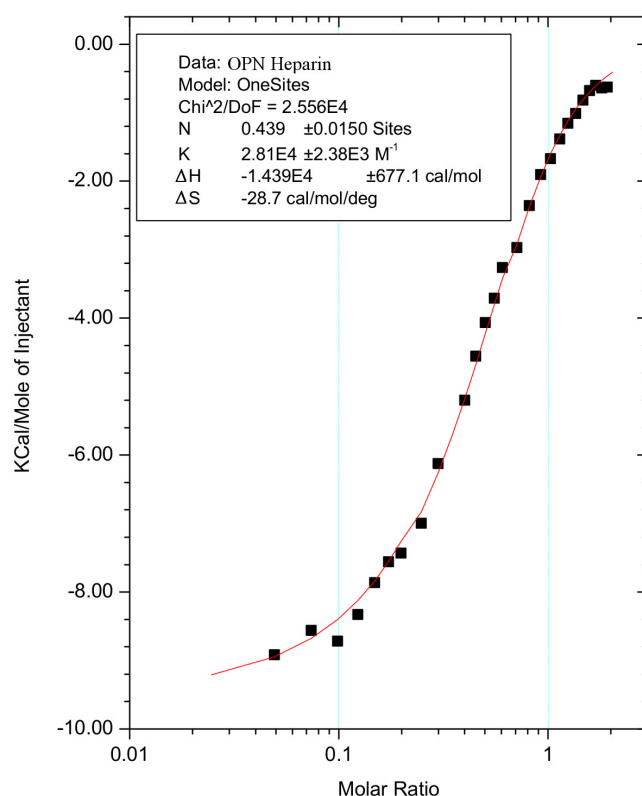


Figure S4. ITC least squares fit of incremental heat release per mole of ligand from 29 automatic injections of Heparin (1.5mM) into OPN (135μM). The data in the inset represents a single run (as opposed to the average of 3 given in the main text). Note that the molar ratio is plotted on a logarithmic scale in order to make the rational for the injection scheme easily visible.

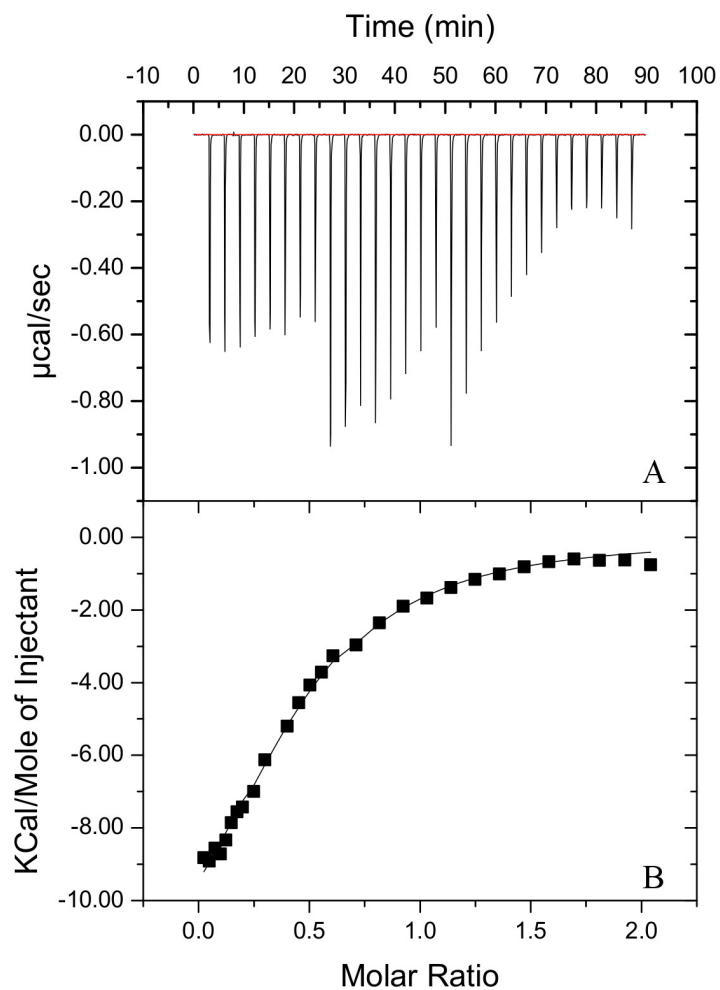


Figure S5. a) Raw data from 29 automatic injections of Heparin (1.5 mM) to an OPN solution (135 μM). b) Least square fit of incremental heat release per mole of ligand. The obtained stoichiometry indicates 0.42 heparin strands per OPN molecule.

Experiments Under High Salt Conditions

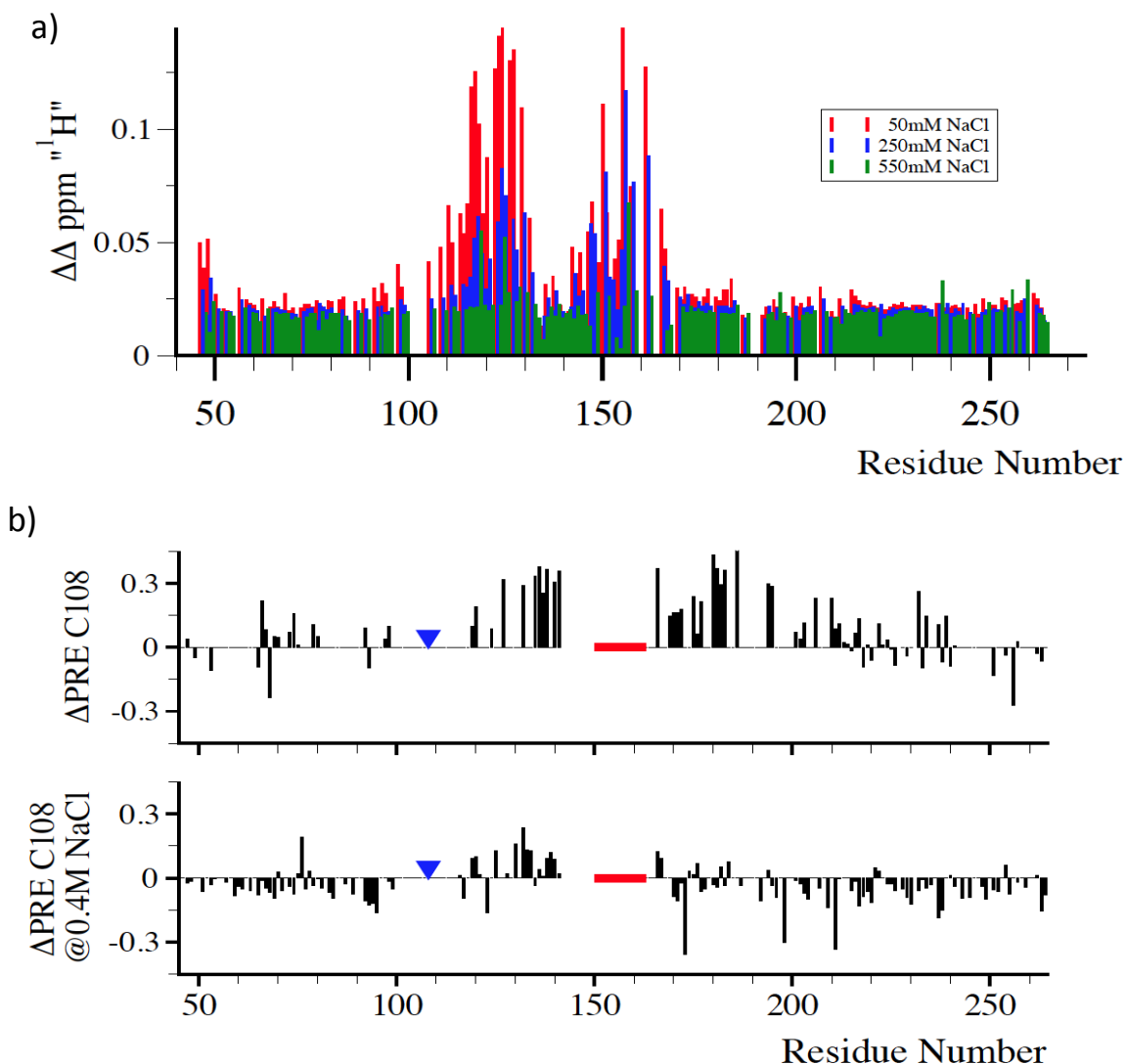


Figure S6. a) Residue plots of ΔCS : Chemical shift change of OPN ^{15}N - ^1H backbone resonances upon addition of heparin at salt concentrations of 50mM (red), 250mM (blue) and 550mM NaCl (green) (nitrogen ΔCS was divided by a factor of 5 before being combined with the proton shift to yield a “pseudo proton shift”). b) ΔPREs : Changes in PRE intensity of mutant C108 induced through the addition of heparin at a concentration of 50mM (top) and 400mM NaCl (bottom). All plots indicate that the CS and PRE effects induced by heparin diminish in the presence of high salt concentrations. Since NaCl nearly exclusively screens electrostatic interactions, these data hint towards a primarily electrostatically governed interaction mode between heparin and OPN. The red bars indicate residues that were excluded from the analysis due to significant line broadening.

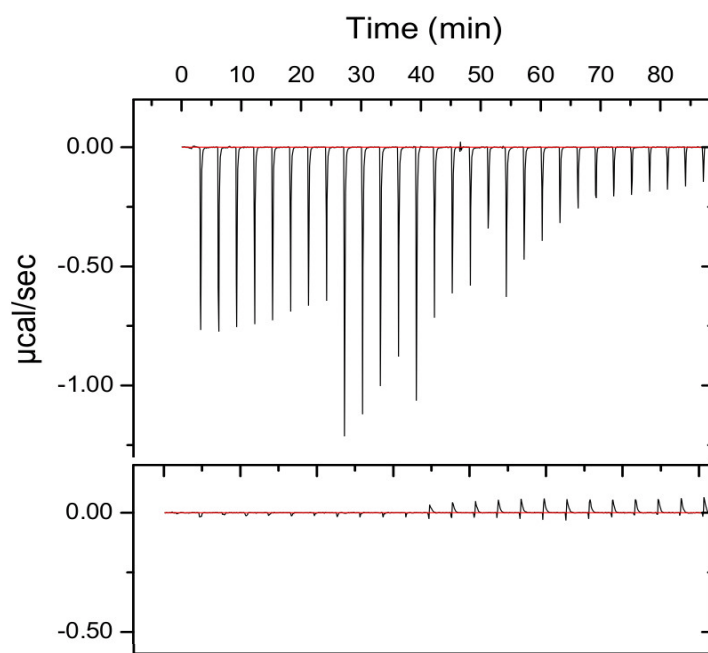


Figure S7. ITC at high Salt Concentrations. a) Raw-data profile for Heparin (1.5 mM) titration into OPN (0.12 mM) at 50 mM NaCl. b) Raw-data obtained for a titration with the same injection scheme, temperature and similar concentrations (0.85 mM Heparin, 0.1 mM OPN) at a salt concentration of 400 mM NaCl. The data was not processed as a different ensemble of OPN would be investigated, thus making comparison of the results very complicated and unfeasible. (Samples were dialyzed into high salt buffer for 48 h. Control-titrations of Heparin into buffer and buffer into OPN were run to test the data quality.)

Error Estimation of PRE Data

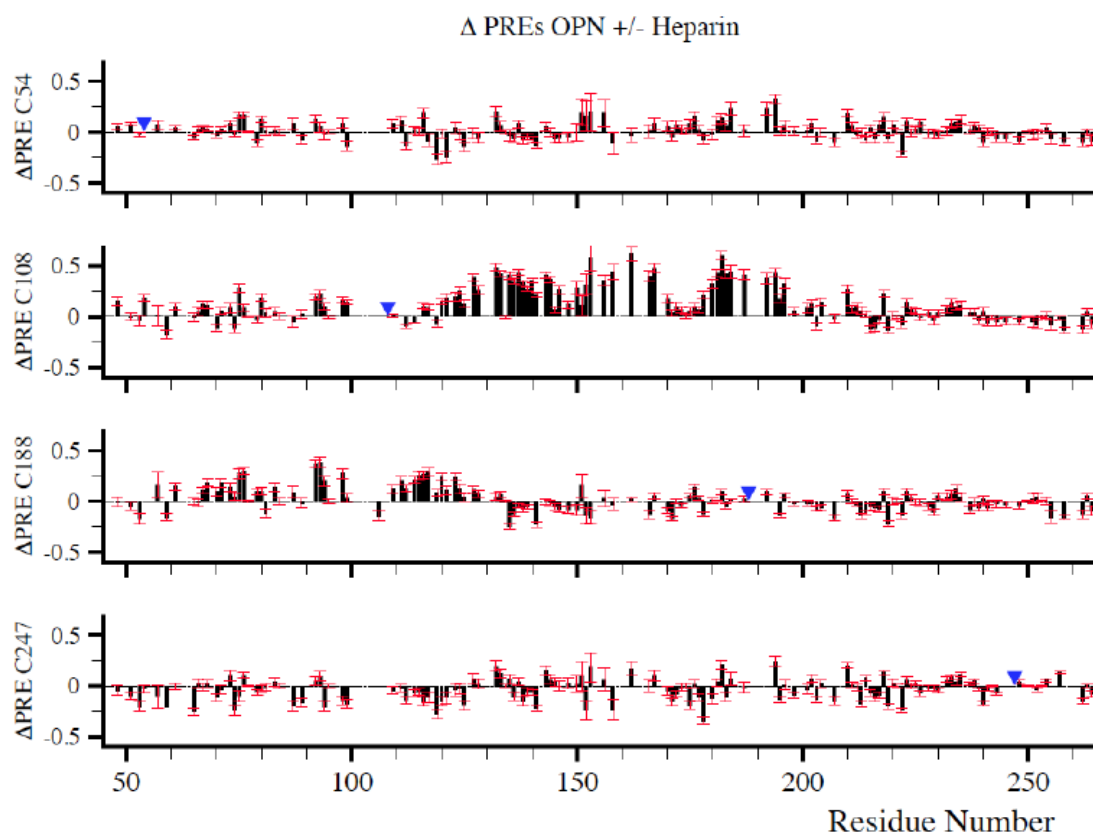


Figure S8. PRE data of the four spin labeled OPN mutants including experimental error estimations (shown in red).

References

- [1] G. Platzer, A. Schedbauer, A. Chemeli, P. Ozdowy, N. Coudeville, R. Auer, G. Kontaxis, M. Hartl, A. J. Miles, B. A. Wallace, O. Glaser, K. Bister, R. Konrat, *Biochemistry* **2011**, *50*, 6113-6124.
- [2] M. J. N. Junk, H. W. Spiess, D. Hinderberger, *Angew. Chem. Int. Edit.* **2010**, *49*, 8755-8759.
- [3] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, L. E. Kay, *Biochemistry* **1994**, *33*, 5984-6003.
- [4] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277-293.
- [5] T. D. Goddard, D. G. Kneller, University of California, San Francisco.
- [6] T. Wiseman, S. Williston, J. F. Brandts, L. N. Lin, *Anal. Biochem.* **1989**, *179*, 131-137.
- [7] N. Origin (OriginLab, MA).

Chapter 3

Magnetic Resonance Access to Transiently Formed Protein Complexes

Sara, T., Schwarz, T.C., Kurzbach, D., Wunderlich, C.H., Kreutz, C., Konrat, R.
Chemistry Open **3**, 115-123 (2014)

ABSTRACT

Protein-protein interactions are of utmost importance to an understanding of biological phenomena since non-covalent and therefore reversible couplings between basic proteins leads to the formation of complex regulatory and adaptive molecular systems. Such systems are capable of maintaining their integrity and respond to external stimuli, processes intimately related to living organisms. These interactions, however, span a wide range of dissociation constants, from sub-nanomolar affinities in tight complexes to high micromolar or even millimolar affinities in weak, transiently formed protein complexes. Herein, we demonstrate how novel NMR and EPR techniques can be used for the characterization of weak protein-protein (ligand) complexes. Applications to intrinsically disordered proteins and transiently formed protein complexes illustrate the potential of these novel techniques to study hitherto unobserved (and unobservable) higher-order structures of proteins.

Introduction

Protein-protein interactions are fundamental chemical events in living organisms and thus of prime interest in many fields of life science. Although protein X-ray crystallography has delivered an enormous amount of highly refined structural data of large protein aggregates its applicability is limited to high-affinity and stably structured protein complexes. In the recent past, it has yet become clear that also weak protein-protein interactions leading to transiently formed complexes play key roles in diverse biological processes, such as cell signalling, post-translational modifications, general transport phenomena, cellular trafficking, enzyme catalysis, and transcription as well as translational regulation. Additionally, there is growing awareness of the importance of intrinsically disordered proteins (IDPs) in eukaryotic life. Their central role in protein interaction networks has been established. IDPs efficiently sample a vast and heterogeneous conformational space allowing them to interact with multiple and diverse binding partners. It is therefore a great and rewarding challenge for modern structural biology research to access these proteins in the completeness of their native states and to adequately describe their inherent structural plasticity. In this

context it is of interest that sparsely populated protein conformational states (high-energy, excited states) are essential for the functionality of proteins – ranging from IDPs to tightly structured enzymes.¹⁻³ It is thus mandated that the conventional notion “protein structure determines function” needs to be reinterpreted (up to now predominantly ground state structures were considered).

Evidently disordered systems are not amenable to X-ray crystallography, which is by far the most widespread structural biology technique. Instead these systems require alternative technological strategies. Nuclear magnetic resonance (NMR) spectroscopy offers unique possibilities for this quest. Because NMR is performed in solution there are different sample requirements. As such, it can also be applied to molecular systems displaying substantial degrees of conformational flexibility. Over the years, NMR methodologies were developed to investigate protein complexes. NMR chemical shift changes and saturation techniques are used to locate protein interaction sites. Residual dipolar couplings (RDC)^{4,5} together with nuclear-Overhauser-effect (NOE)-based methods⁶ and paramagnetic relaxation enhancement (PRE)⁷ can be used to evaluate binding interfaces with atomic resolution, even in cases of weak affinities and transient intermolecular contacts. It should be noted that NMR not only allows locating interaction sites and provides structural information about the binding interface, but also allows for the quantification of binding affinities (dissociation constants, K_D). For more details and broader coverage of NMR applications, the reader is referred to excellent reviews published over the years.^{6,8,9} Here we provide an illustration of the power of NMR spectroscopy for the elucidation of reversible protein interaction events by means of so-called relaxation dispersion (Car-Purcell-Meiboom-Gill, CPMG) experiments. Applications to studies of protein-protein interaction are illustrated with a quantitative analysis of Max homodimerization. It is shown that NMR provides detailed information about this dynamic protein dimerization process. Both kinetic and structural parameters are obtained, which permits the analysis of the only marginally populated protein complex. An example for studies of enzyme catalysis is given through the observation of reversible binding of the cold-shock protein to its cognate cold-box RNA. Here, dynamics in unperturbed equilibrium of both, the protein and RNA components, can be studied via NMR spectroscopy. Again, we anticipate that these NMR measurements will provide unprecedented insight into the molecular details and elementary steps of

RNA chaperone function. Finally, we present a detailed description of how NMR and electron paramagnetic resonance (EPR) spectroscopy can be used in a complementary fashion to investigate complex formation of IDPs. Most importantly, this novel approach revealed that upon ligand binding the IDP Osteopontin largely remains disordered although its conformational ensemble is updated. Ligand binding proceeds through conformational selection of a preformed conformational substate that is characterized by an unexpected cooperative stabilization as common for stably folded globular proteins. The observed structural and dynamical compensations upon binding reveal how thermodynamically unfavorable entropic penalties are compensated through partial unfolding of peptide segments remote to the primary binding site.

Results and Discussion

Access to sparsely populated conformational states by relaxation dispersion NMR experiments

Biomolecules are characterized by rugged energy surfaces resulting from attractive and repulsive forces of similar strength. The accessible conformational space governs thermodynamics and kinetics of conformational transitions between different substates. In the past it has become increasingly evident that sparsely populated, high-energy conformational states of a protein can be essential for protein function.^{3,10} A prominent example is the frequently observed conformational selection mechanism for substrate recognition.^{11,12} In this context a substrate is assumed to interact with a (sparsely) populated high-energy (excited) conformation of a protein and to modulate the system's conformational ensemble by stabilizing this particular high-energy state. However, excited conformational states are difficult to access by means of conventional (direct detection) NMR, since fast conformational averaging leads to averaging of resonances. Averaging “hides” a sparsely populated state behind predominating low-energy states. Consider a conformational transition process in which a highly populated (low energy) ground state exchanges with a lowly populated (high energy) excited state. The resulting NMR signal is a population weighted average of the individual signals stemming from the two states and dominated by the abundant ground state signal. Detailed information about the sparsely populated excited state can nevertheless be obtained, by the so-called CPMG (Car-Purcell-

Meiboom-Gill) relaxation dispersion experiment.^{13,14} This experiment takes advantage of the fact that the conformational transitions between a highly (A) and a lowly (B) populated species affects the transverse relaxation times and thus line widths and intensities of ground states' (A) NMR signals. Suppose a 90° pulse that flips the magnetization of an amide nitrogen of a particular amino acid from the quantification axes to the transverse plane. The spin will immediately start to precess and concomitantly dephase during a given time interval τ . Conformational transitions between ground and excited state lead to alterations of the precession frequencies (Fig.1a, top). Application of a 180° inversion pulse after the time interval τ leads to partial refocusing of the precessing magnetization and hence increased signal intensities after the interval 2τ (Fig 1a. middle). Increasing the number of 180° inversion pulses improves the refocusing efficiency in the CPMG experiment and results in increased signal intensities (Fig.1a, bottom). Depending on the overall exchange rate between the two states A and B, $k = k_A + k_B$, (with k_A and k_B being the forward and backward rates) the number of refocusing pulses will alter the observed signal intensity. If the exchange rate k is slower than $1/\tau$, the precessing magnetization will be completely refocused and signal intensity is unaffected by the exchange process and hence maximized. Yet, if k is in the range of or larger than $1/\tau$ the observed signal intensity will be modified and depend on the frequency (ν_{CPMG}) of 180° inversion (refocusing) pulses. In other words, the effective (detected) transverse relaxation rate, $R_{2,\text{eff}}$ will be affected by the exchange process and the time interval τ (or the number of 180° inversion pulses, see Figure 1). Increasing the frequency of 180° inversion pulses, where $\nu_{\text{CPMG}} = 1/(4\tau)$, in the CPMG train (Figure 1a) will lead to progressive quenching of the exchange process. This dependence is exploited by CPMG pulse sequences in which pulse trains with repeating $\tau - 180^\circ_x - \tau$ blocks are applied to refocus the magnetization. Sampling different values of ν_{CPMG} yields a dependence of $R_{2,\text{eff}}$ *versus* ν_{CPMG} . $R_{2,\text{eff}}(\nu_{\text{CPMG}})$ will decrease with increasing ν_{CPMG} values (Figure 1b). In the case of complete quenching of the exchange process, however, one observes a flat profile with $R_{2,\text{eff}} = R_{2,0}$ ($R_{2,0}$ is the exchange-free transverse relaxation rate). Fitting $R_{2,\text{eff}}(\nu_{\text{CPMG}})$ as a function of ν_{CPMG} allows to determine the exchange rate k between the two states. Furthermore, the chemical shift difference between A and B and the populations of these states can be extracted through measurements at different magnetic field strengths.¹⁵ Depending on the time

scales, different approaches have to be taken into account for quantitative analysis of fast, slow and intermediate exchange regimes.¹⁶

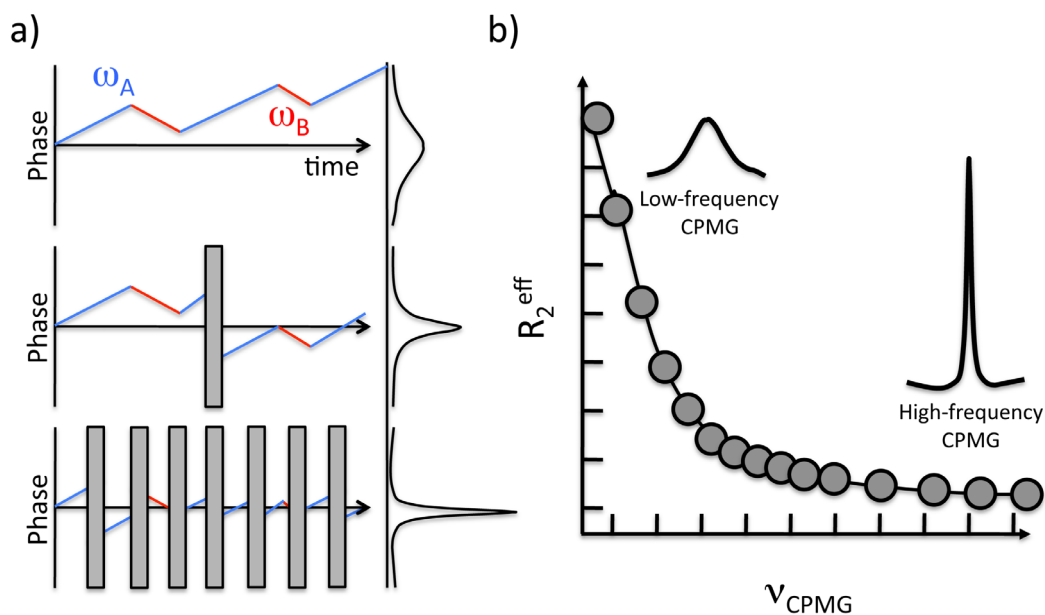


Figure 1. a) Phase of two spins with resonance frequencies ω_A (blue) and ω_B (red) in dependence of pulse sequence evolution time and pulse interlay. Corresponding detectable NMR signals on the right. With increasing ν_{CPMG} the signal becomes sharper. b) Theoretically calculated dependence of $R_{2,\text{eff}}$ on ν_{CPMG} .

The CPMG experiment can also be explained from a kinetic point of view by taking into account the deleterious effects of randomly changing precession frequencies on echo formation. The observed NMR signal, that is, a spin echo, will become weaker the more the precession frequencies of the individual spins in an ensemble differ from each other. The time needed for the spins to be randomly distributed in the transverse plane after RF excitation is inversely proportional to the spectrum of precession frequencies (line width). Conformational exchange modulation of precession frequencies due to changes in the local environment of some spins leads to additional line broadening and concomitant signal attenuation. Taking a look at the CPMG pulse train in Figure 1a one can see that this effect will only be effective if the time between two pulses is longer than the spin needs (on average) to change its precession frequency due to conformational changes in the system. In other words, the frequency of refocusing pulses (ν_{CPMG}) must be shorter than the exchange rate k in order to observe enhanced signal attenuation. Thus, by varying ν_{CPMG} and observing

the consequent effect on the signal intensity one can yield information about k and the details of conformational exchange processes (frequency difference between individual states).

In the realm of protein NMR one frequently observes the relaxation dispersion of backbone ^{15}N - ^1H amide resonances. Implementation of the CPMG train into a two dimensional data acquisition yields residue dependent $R_{2,\text{eff}}$ values. In this case a residue's exchange frequency might be affected by local conformational interconversion processes, for example, flapping of aromatic side chains,¹⁷ and by global processes like large scale conformational adaptations due to, for example, ligand binding. In order to separate global from local influences one can choose to fit the dispersion profiles in a global or residue-specific manner. Numerous applications have demonstrated the wide applicability of this approach and provided valuable insight into the atomic details of relevant dynamic processes in biology.¹⁸ Here we illustrate the technique with applications to a transiently formed chaperone-RNA complex and the reversible formation of a homodimeric coiled-coil protein complex.

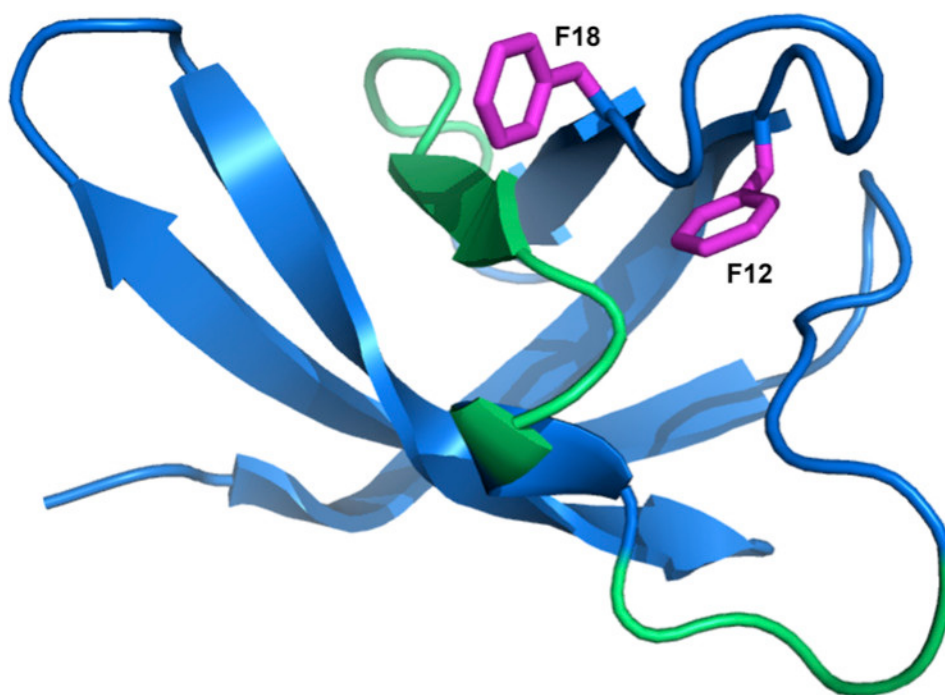


Figure 2. Structural model of CspA with surface hydrophobic residues shown in magenta; the primary RNA binding site is represented in green (residues 26-41). Structure PDB ID: 2L15.¹⁹

As a case study of lowly populated state of ligand-bound protein we present here the interaction of the *Escherichia coli* cold shock protein CspA (the 3D structure of the apo-state of CspA is given in Figure 2) with a fragment of its mRNA. The CspA is a 70 amino acid 7.4 kDa protein naturally occurring in cold-shocked bacteria. It functions as an RNA chaperone by interacting with the 5'-untranslated region of its own mRNA (called the coldbox RNA, CB RNA);²⁰ overproduction of the CspA mRNA causes derepression of the cold-shock genes expression. The CspA structure comprises a β -barrel fold. The RNA binding surface is primarily constituted of solvent-exposed aromatic residues, mostly phenylalanines.^{20,21} However, the RNA bound state of the protein is difficult to access by means of directly detected NMR, since only a fraction of the present RNA molecules interact with CspA at any given time, and the time frame of the interaction is very short. Saturating the population of the bound state is unfeasible due to the progressive signal loss with increasing RNA concentration as a consequence of significant line broadening, especially in the binding sites.²⁰ Most other techniques for measuring protein-ligand interaction are not applicable since the population of the bound state protein is usually below their limit of detection.

Thus, the CPMG relaxation dispersion (RD) experiment is a promising technique to supply information about the sparsely populated RNA bound state of CspA. Here, we provide ¹⁵N-CPMG data of residues F12 and F18 (Figure 3) to illustrate how this technique can be used to analyze the transiently formed CspA-RNA complex. The employed RNA interaction partner is the so-called anti-coldbox (ACB) RNA, which is partially complementary to the cold-box RNA.

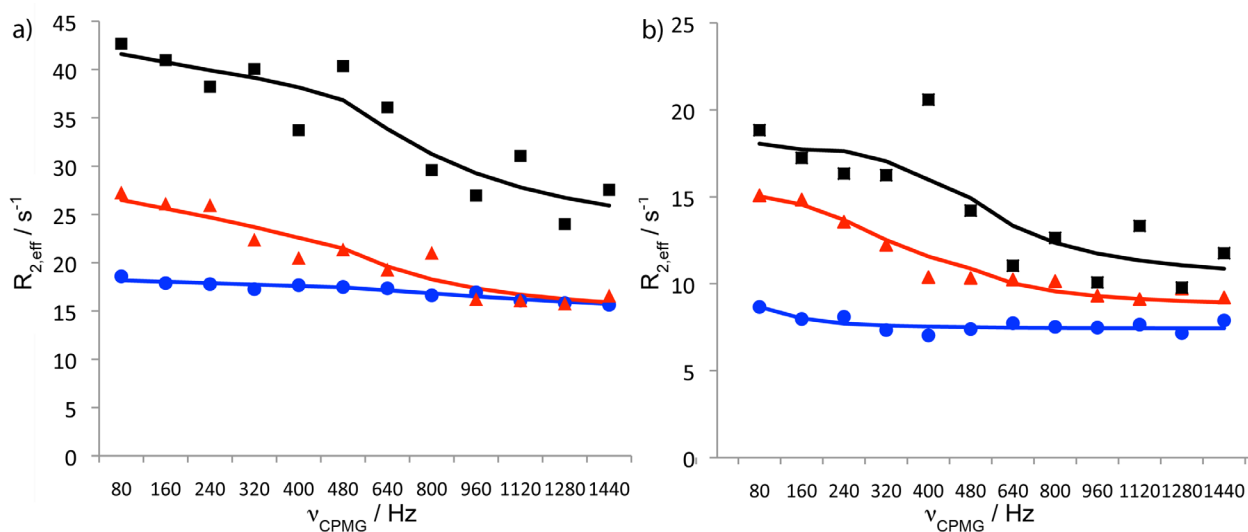


Figure 3. CPMG relaxation dispersion profiles for CspA residues a) F12 and b) F18. Relaxation profiles of the CspA in the presence of 10% ACB (▲ and ■) and in the absence of RNA (●). Black squares (■) show the relaxation profile at 800 MHz NMR field frequency and red triangles (▲) show the profile measured at the frequency of 500 MHz; solid lines indicate “global” data fits.

For CspA residues F12 and F18, ^{15}N -dispersion relaxation profiles were measured in the absence and in the presence of ACB-RNA (Figure 3). In the absence of RNA no exchange process was detected as reflected in the flat dispersion relaxation profiles. However, in the presence of 10% molar ratio of ACB-RNA, distinct changes in relaxation profiles were observed. Global fitting based on an intermediate exchange model,¹⁶ yielded 1.6 % of the conformational ensemble to comprise the CspA RNA-bound state; the exchange frequency was determined to be 1.4 kHz. The participation of the phenylalanine residues F12 and F18 (highlighted in Figure 2) adjacent to the primary RNA binding site during the binding event was previously described.²¹ It is noteworthy that the relaxation dispersion profiles do not necessarily reflect the kinetics of the RNA binding event, but the frequency of the structural rearrangements in the region comprising residues F12 and F18. However, it has been shown that the RNA binding triggers structural adaption of the protein. Hence, it is reasonable to correlate the exchange frequency (1.4 kHz) fitted globally to F12 and F18 with the effective time constant of the RNA binding/release. Referring to the structural model shown in Figure 2, the F12 and F18 residues are positioned sideways to the primary binding site and can therefore establish stacking interactions with the ribose moieties in the bound

RNA strand. Other residues not present in either the primary or the secondary binding site do not exhibit changes in $R_{2\text{eff}}$ as observed for F12 and F18 (data not shown). Thus, CPMG analysis allows for determination of kinetics of a state as lowly populated as 1.6 %, which would be difficult to assess by other techniques. Further, the method used here enables one to obtain information not only on kinetics and population, but also yields the nitrogen chemical shift differences between the free and the bound state of residues observed by CPMG data that correlate with the magnitude of changes in chemical environment. Thus, in principle one could extract structural information about the bound state of CspA by means of CPMG through the chemical shift difference to the apo-state if a sufficiently high number of residues yield exchange-affected dispersion profiles. The CspA case study presented here demonstrates the power and versatility of the CPMG method in assessing the lowly populated states important in RNA-protein interactions.

To complement the results obtained from CspA-derived CPMG we also focused on the investigation of micro- to millisecond dynamic processes of the CB and ACB RNAs, which interact with CspA. In contrast to protein NMR spectroscopy, where the backbone amide nitrogen-proton spin pairs represent suitable reporter nuclei for CPMG RD experiments, for nucleic acids, ^{13}C -CPMG RD experiments are the method of choice.²² This is due to the fast exchange rates of the imino protons, that is, nitrogen bound protons (H3 in uridine and H1 in guanosine) directly involved in base pairing, with bulk water leading to significant line-broadening effects or even to a total collapse of the imino proton signal within the dominant water resonance. But as in protein NMR spectroscopy, uniform $^{13}\text{C}/^{15}\text{N}$ -labeling patterns impair the application of certain NMR experiments to address dynamic features, like the CPMG RD experiment. A uniform $^{13}\text{C}/^{15}\text{N}$ -labeling pattern introduces one-bond scalar and dipolar ^{13}C - ^{13}C couplings resulting in resolution and sensitivity problems especially for larger RNA comprising more than 30 nucleotides (nt) but also to artifacts in the CPMG RD data sets impairing a reliable analysis.²³ To circumvent several issues arising from the state-of-the-art uniform labeling protocol using $^{13}\text{C}/^{15}\text{N}$ -modified ribonucleotide triphosphates (rNTPs) and T7 RNA polymerase, several approaches were proposed.²⁴ In one of our laboratories, we have recently put focus on isotope labeling RNAs via the solid phase synthesis approach. A minimally invasive ^{19}F -based isotope labeling protocol can be very efficiently applied to RNA as all four 2'-fluorine modified

nucleotide phosphoramidites are commercially available and can be incorporated into the target RNA without modifying the standard protocol (Figure 4a). The ^{19}F nucleus has favorable NMR spectroscopic features, like 100% natural abundance, an intrinsic NMR sensitivity almost as high as protons (83%) and a vast chemical shift range about 100 times larger than protons. It represents a bioorthogonal reporter spin for nucleic acids and was efficiently used in RNA folding and RNA ligand binding NMR assays.²⁵⁻²⁷ We currently introduce 2'-fluorine labels into the CB and ACB RNAs to study the interaction of both RNAs with CspA but also to address conformational dynamics via ^{19}F -CPMG relaxation dispersion experiments (data not shown).

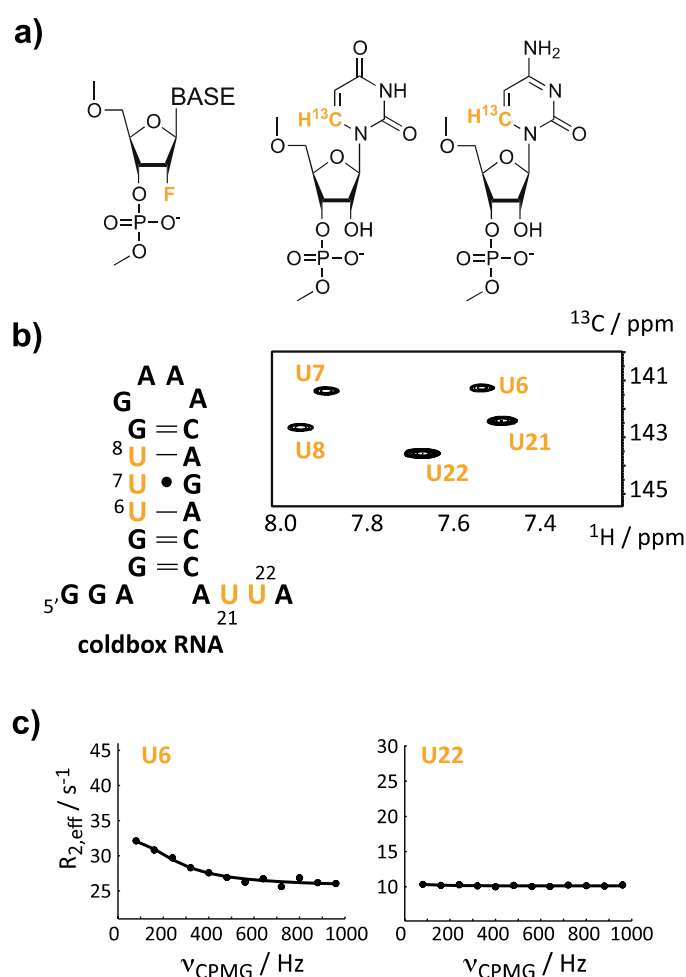


Figure 4. a) Stable isotope labeling patterns that can be introduced via the solid phase RNA synthesis approach. 2'-Fluorine and 6- ^{13}C -pyrimidine labels are shown. b) Coldbox RNA with five 6- ^{13}C -uridine labels and the ^1H - ^{13}C -HSQC spectrum with assignments. c) Selected ^{13}C -CPMG relaxation dispersion profiles at 125 MHz carbon larmor frequency of the coldbox RNA. Residue U6 shows a non-flat dispersion profile indicating microsecond to millisecond dynamics, whereas U22 displays no dynamics on that timescale.

We also focused on the chemical synthesis of site-specifically ^{13}C -modified pyrimidine phosphoramidites (Figure 4a).²⁸ This labeling protocol results in functionally unperturbed RNAs at the costs of the labor-intensive production of the 6- ^{13}C -uridine and -cytidine phosphoramidites. For example, all uridines in the aforementioned stem-loop motif (CB RNA, Figure 4b) residing in the 5'-UTR of the *E. coli cspA* messenger RNA (mRNA) were replaced by the 6- ^{13}C -modified counterparts. We then used ^{13}C -CPMG relaxation dispersion experiments to address the micro- to millisecond dynamics of this very RNA. Preliminary results are shown (Figure 4c) and hint towards a dynamic hotspot of the coldbox RNA in the vicinity of the G•U wobble base pair. A quantitative description of this dynamic phenomenon and the possible biological relevance for the interaction with CspA are currently investigated in our research groups.

As a third example for the application of CPMG techniques to transient protein interaction studies, we outline an application to the reversible protein dimerization of the Myc associated factor X (MAX).²⁹ MAX readily forms heterodimers with its authentic binding partner Myc yielding an active transcription factor complex. Deregulation of this tightly controlled protein complex formation can lead to severe disease phenotypes.²⁹ Under neutral pH-conditions and room temperature MAX natively populates predominantly a dimeric, coiled-coil α -helical motif (see Figure 5a).³⁰ Changing solution conditions such as pH and temperature, however, allows for controlled manipulation of the monomer-dimer equilibrium.³⁰ In Figure 5b) a ^{15}N - ^1H HSQC of MAX is shown under conditions that favour the monomeric form (pH 5.5, 35°C).³⁰ The strong spectral overlap indicates that the prevailing monomeric form of MAX is to a large degree conformationally disordered. In Figure 5 c,d we show CPMG profiles of two residues, Ser32 and Glu83, of MAX under these conditions at 500 and 800 MHz NMR static field strength. The observed exchange process corresponds to interconversion of monomer and dimer MAX states.³⁰ Fitting the data globally with an intermediate exchange model yields a population of the Max dimer of 0.75 % with an exchange rate of 0.4 KHz.¹⁶ This is in excellent agreement with data obtained in a thermodynamic analysis of the Max monomer dimer equilibrium.³⁰

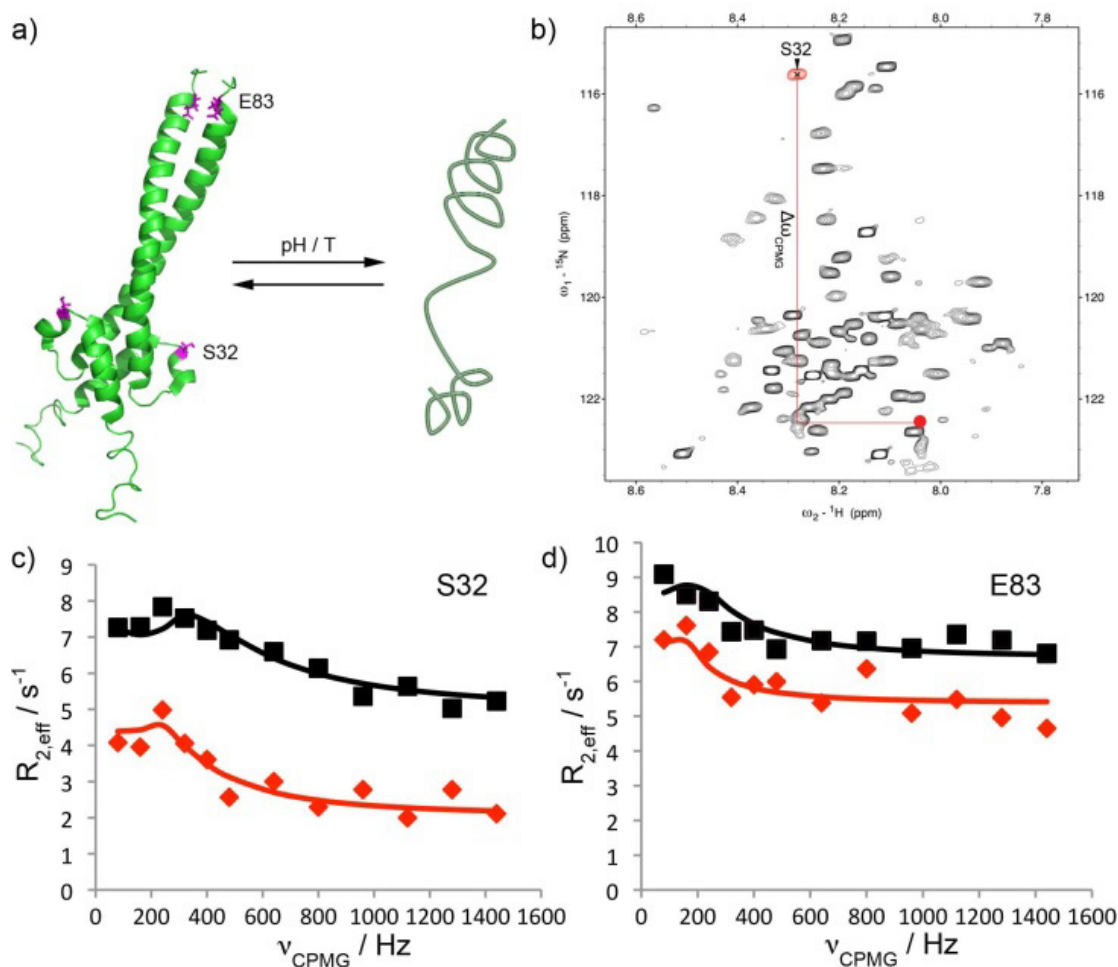


Figure 5. a) Sketch of the dimerization equilibrium of MAX. b) ^{15}N - ^1H HSQC of the monomeric state of MAX. The red dot indicates ^{15}N and ^1H frequencies for residue S32 in the homodimeric coiled-coil state (taken from Ref.[29]). c,d) CPMG relaxation dispersion profiles at 500 (red) and 800 (black) MHz. c) Residue Ser 32 d) Residue Glu 83. The solid lines represent fits of the data with an intermediate exchange model.

The reliability of the CPMG results can be further tested by comparison of the chemical shifts extracted from the CPMG data with literature data for the dimeric state of MAX. From literature, for example, the ω_{N} of the bound (dimeric) state of Max is known, while the chemical shift of the monomeric form can be directly observed in ^{15}N - ^1H HSQC spectra (since the monomeric form of Max is the dominating species under these conditions). The differences between monomeric and dimeric chemical shifts should coincide with individual $\delta\omega$'s obtained from a global fit of CPMG data. As indicated in Figure 5b, for Ser 32 excellent agreement between fitted and experimental chemical shift differences were found ($\delta\omega_{\text{fit}} = 8.0$ ppm and $\omega_{\text{N,Dimer}} - \omega_{\text{N,Monomer}} = 7.8$ ppm).²⁹ Since the chemical shift change depends on molecular

rearrangements (e.g., through changes in conformation, local hydration, hydrogen bonds etc.), unique information can be obtained about the atomic details of sparsely populated protein states even under conditions where other experimental techniques fail as a consequence of the low population.

Assessing IDP sub states by means of combined EPR and NMR

The hallmark of intrinsically disordered proteins (IDPs) is their vast and heterogeneous conformational ensemble comprising both extended conformations and simultaneously compact reasonably stable structures displaying significant resistance against denaturing agents (urea).³¹ The rugged energy surface typically found for IDPs allows for conformational transitions between different sub states and endows IDPs to transiently interact with a multitude of binding partners via, for example, conformation selection type processes. Despite their enormous biological relevance, the structural and dynamical characterization of this important protein family is still far from routine. Most importantly, the existence of sparsely populated sub states in the conformational ensemble requires highly sensitive experimental techniques to probe their structural features and encounters with protein binding partners. It was recently demonstrated that electron paramagnetic resonance (EPR) spectroscopy is a very powerful experimental tool to access sparsely populated states and transient protein interactions.³²⁻³⁵ Furthermore, EPR spectroscopy is a promising candidate to supplement NMR data in terms of time and length scales. Some pilot studies of EPR on IDPs have already been published,³²⁻³⁵ yet the strength of EPR, i.e., the detection of through-space dipolar interaction of electron spins by double electron-electron resonance (DEER) spectroscopy,³⁶⁻³⁹ has turned out to be difficult to apply to IDPs. The commonly applied analysis techniques for DEER raw data fail in the case of IDP conformational ensembles.⁴⁰ Therefore, we have amended the repertoire of analysis methods for DEER data with an approach that is suitable to interpret data obtained for IDPs comprising large and heterogeneous conformational ensembles. In the following section, we will briefly introduce the standard DEER methodology and explain our expansion of this method to IDPs. Our investigations based on this novel approach are accompanied by complementary NMR experiments. Generally speaking, we showed that NMR and EPR data supplement each other and provide a comprehensive picture of IDPs' structural dynamics.

In order to perform DEER, one has to modify a protein with two labels that both carry an unpaired electron, hence the term spin label.⁴¹ The free electrons of the two labels will experience a through space dipolar coupling. However, since at room temperature dipolar couplings between two labels in solution average to zero the sample has to be freeze-quenched. As such, one measures intramolecular the dipolar couplings between every pair of spin labels “locked in” at the glass transition temperature (after correcting for intermolecular background contributions). In order to prolong relaxation times, DEER is, yet, measured at even lower temperatures, typically between 20 and 50K. The dipolar couplings manifest themselves as modulation of DEER time domain data, that is, the signal intensity as a function of the pulse sequence evolution time.^{31,38,42} For a single, fixed distance, one would get a single cosine modulation of the time trace. Yet, as the conformational ensemble grows, the number of populated substates increases and each substate contributes to the DEER time trace. The frequency of the cosine modulation is dependent on the interspin distance. For the common DEER analysis methods, the individual contributions of the different substates are given primarily by their weighted populations (note that average interspin distances can become of importance in large disordered systems, too).³¹ With increasing number of populated conformations and corresponding interspin distances, the total, observable modulations will become increasingly blurred and will ultimately converge to an exponential decay. This circumstance gives rise to particular problems in separating intra- and intermolecular contributions to the DEER signal for broad distributions of distances. In the case of doubly labelled proteins, one typically wants to eliminate intermolecular contributions. For intrinsically disordered systems, this can only be achieved by measuring DEER references on singly spin-labelled proteins, which yields the pure intermolecular contributions.⁴¹

In case of folded proteins, after removal of background contributions, the spin label distance distribution $P(R)$ can be obtained by well-established procedures such as Thikonov regularization.^{41,43} Yet, in the case of IDPs the broad and inhomogeneous distance distributions leads to less significant time trace modulations and poor signal-to-noise ratios that impairs the extraction of feasible distance distributions (e.g., like the traces shown in Figure 6).⁴⁰ Thus, we proposed an alternative approach based on a so-called effective modulation depth, Δ_{eff} , for data interpretation for IDPs. Details of this novel approach were explained elsewhere.³¹ In short, Δ_{eff} denotes the signal decay

of the time trace, $V(t)$, at a given point of time, t_{eff} . Δ_{eff} can be defined as: $\Delta_{\text{eff}} = 1 - V(t_{\text{eff}})/V(t=0)$. For large IDPs, we suggest to simply choose the longest experimentally achievable DEER evolution time for t_{eff} . Δ_{eff} is an approximate measure of the average interspin distance for broad $P(R)$ s. Yet, the reader should be aware that Δ_{eff} does not linearly depend on the population-weighted average distance in the measured ensemble but is a complex function of several spectroscopic parameters.³¹ To a first approximation, however, Δ_{eff} decreases with increasing interspin distance R for broad distance distributions. With Δ_{eff} as tool in hand, we probed structural preferences of Osteopontin (OPN), a cytokine involved in metastasis of several kinds of cancer.⁴⁴ Typical DEER data obtained on an IDP are shown in Figure 6.

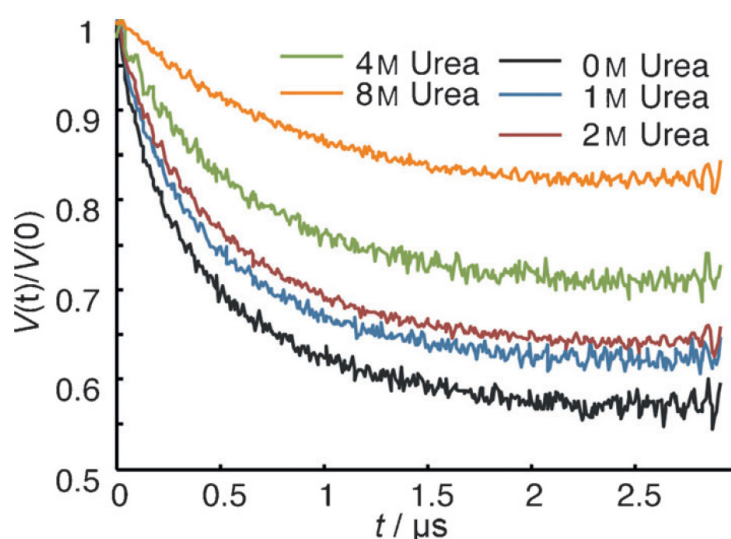


Figure 6. DEER time traces of a double mutant comprising the central OPN segment from residue C108 to C188 at different urea concentrations.

Conformational stabilities, understood as resistance to urea unfolding, of several individual structural segments of OPN were investigated by recording DEER time traces for different spin labelled double mutants in dependence of increasing urea concentration. As such, a decrease of Δ_{eff} with increasing urea concentration is representative for unfolding and expansion of a doubly spin labelled protein of interest. In Figure 6 exemplarily experimental Δ_{eff} values are shown as a function of urea concentration for a selected double mutant comprising a central segment of the cytokine OPN. The time traces appear as exponential decays because of the aforementioned convergence of cosine functions (one for each conformation in the OPN ensemble). With increasing urea concentration Δ_{eff} decreases indicating an expansion of the IDP. As published earlier³¹ the C-terminal part of OPN exhibits an

exponential decay of Δ_{eff} with increasing urea concentration. This gives rise to a steep slope of the Δ_{eff} function and can be regarded as a denaturation profile of an unstably folded protein segments of potentially random-coil like character. Already for low urea concentrations such segments show significant conformational expansion (i.e., a decrease in Δ_{eff}) in accordance with the idea of very low stability of transient or residual structural elements in IDPs. For other mutants one might observe an approximately linear decrease of Δ_{eff} with urea concentration, indicating that the OPN segment framed by these mutants is on average conformationally more stable than the C-terminal segment. Nevertheless, it is still largely unstructured, random coil- or (pre)molten globule-like. Strikingly, however, for a mutant of OPN with two terminal labels, we observe a sigmoidal development of the Δ_{eff} -derived denaturation profiles with urea concentration (see Figure 7). Sigmoidality is a hallmark for cooperative folding of protein conformations and unexpected for an IDP.⁴⁵

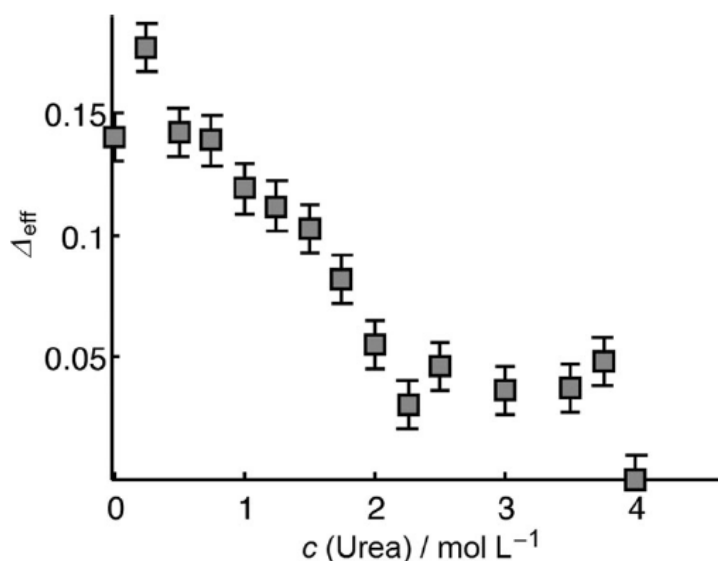


Figure 7. Δ_{eff} for an OPN mutant (C54-C247) comprising nearly the whole protein as a function of urea concentration. Error bars stem from signal noise.

A sigmoidal denaturation profile is indicative for stably and cooperatively folded tertiary structures of OPN, since for low urea concentrations up to 0.75 M the whole protein does not expand significantly (as seen in nearly constant Δ_{eff} values). This observation of a cooperatively folded conformation is surprising as distance distributions between two labelled residues of OPN are generally quite broad, as deducible from prior studies concerning OPN's conformational space¹¹ and as

reflected in the non-modulated DEER time traces (Figure 6). This interesting finding can, however, be understood by concluding that the structural ensemble of OPN contains both, cooperatively folded and unfolded conformations and that both contribute to the DEER signals.³¹ This deduction is possible here only because DEER EPR on freeze-quenched solutions elucidates the whole set of co-existing conformations; ensemble averaged data here would not allow for discerning between partial structuring and sampling of compact conformations. Most importantly, the existence of structural cooperative transitions from folded to unfolded states and vice versa as monitored by EPR in combination with NMR in IDPs calls for a novel conceptual view of IDPs that goes beyond the traditional binary scheme of order vs. disorder. The subtleties of heterogeneous conformational sampling in IDPs and their putative relevance for biological functions have to be adequately addressed.

The applicability of this novel approach to the observation of transient complex formation of IDPs was also recently demonstrated. Not only can the Δ_{eff} approach be utilized to elucidate structural preferences of IDPs complementary to NMR data, but also ligand interaction can be observed in high detail. We analyzed the interaction of the OPN with heparin,¹¹ a highly sulfated glycosaminoglycan widely used as anticoagulant (see Figure 8a).¹⁰ In a biological context heparin binding to OPN is of interest since it models the OPN- heparan sulfate interaction, which constitutes a crucial cofactor in OPN-CD44 receptor association, a process involved in cell signalling and adhesion.⁴⁶ Additionally, interactions between IDPs and biological polyelectrolytes are quite common,³ and our results might well be applicable to other systems.

Upon binding to heparin, OPN largely remains disordered although its structural ensemble is updated. For several doubly spin labeled mutants of OPN heparin binding leads to a clear decrease in Δ_{eff} indicating longer distances between the labeling sites of the six double mutants and an expansion of the protein upon heparin binding. This is shown in a previous paper.¹⁰ These information from EPR agrees well with information concerning the binding event gained from NMR-based paramagnetic relaxation enhancements (PREs).¹⁰ When interpreting PREs of IDPs one should be aware that due to the rapid conformational sampling of IDPs one observes ensemble averaged PRE data.⁴⁷ Hence, all conclusions drawn from these refer to “average”

conformations. Differential values (bound minus apo state PREs) or $\Delta\text{PRE} > 0$ indicate “on average” increasing distance in the protein complex between labeling site and a residue upon binding, $\Delta\text{PRE} < 0$ the opposite. As can be observed in Figure 8b) OPN displays differential changes of long-range backbone interactions as heparin binds. The central spin labels attached to 108C experiences a displacement from the core region around residue 130-190, whereas the spin label 188C is separated from the region comprising residues 90 and 120. Information on the binding process and its impact on OPNs dynamic behavior can be obtained when comparing the ^{15}N NMR relaxation rates and heteronuclear ^{15}N - ^1H NOEs (hetNOE) of bound and free forms. NMR relaxation reports on motions ranging from picosecond to low-nanosecond timescales. The three measurable NMR relaxation parameters (R_1 , R_2 and hetNOE) have different dependencies on the time scales of motions. While the ^{15}N transverse relaxation rate R_2 reports on nanosecond motions, the heteronuclear ^{15}N - ^1H NOE depends on the efficiency of magnetization transfer from ^{15}N to ^1H in the protein backbone, and which is mostly influenced by very fast (picosecond time scale) dynamics of the N-H bond. Figure 8d shows that upon heparin binding to OPN very fast motions (probed by ^{15}N - ^1H hetNOE) increase in the region located around residue 120. The central region around residue 150, however, shows increased ^{15}N transverse relaxation rates R_2 due to decreased conformational flexibility in the binding cleft. Closer inspection of the charge distribution in OPNs primary sequence reveals that the differential mobility changes upon heparin binding are linked to the electrostatic pattern found in OPN (increased mobility in negatively charged regions and decreased mobility/rigidification in positively charged patches).

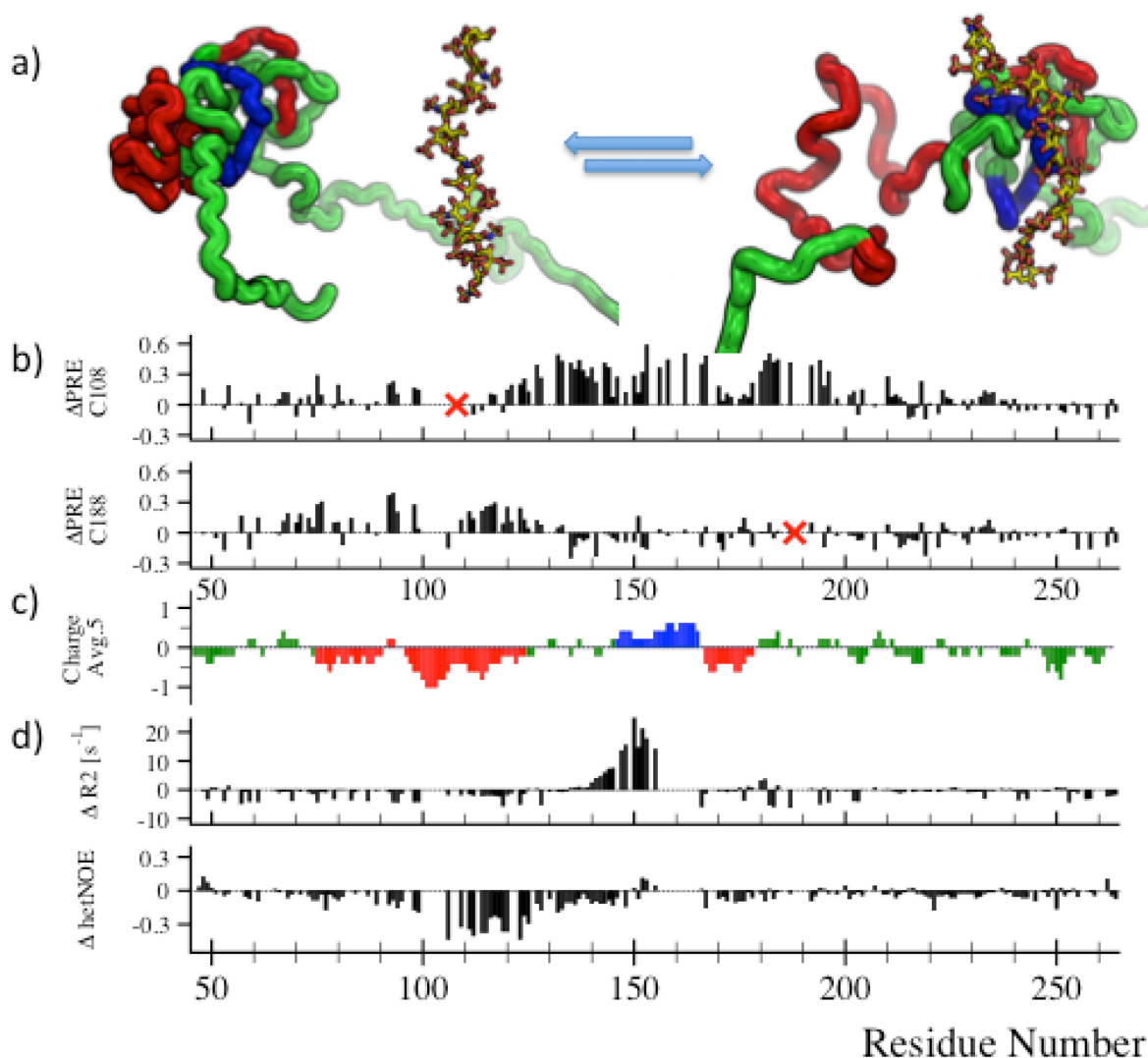


Figure 8. a) Schematic representation of the OPN - heparin interaction. b) Differential PRE values for the two central mutants (PRE of OPN and heparin; PRE of OPN). The positions of spin labels are indicated (red crosses). c) Approximate residue charge of OPN. The color code is identical to the schematic representation of the binding process in Figure 8a. d) Differential transverse relaxation and heteronuclear NOEs (OPN and heparin; OPN).

Through combination of NMR and EPR a clear picture arises of the OPN-heparin binding event. As the highly negatively charged heparin contacts OPNs positively charged core region, the structural and dynamic properties of the OPN ensemble get drastically altered. Along with an increase in dynamics along most of the proteins backbone, an average increase in distances is observed. This unfolding-upon-binding event is depicted in Figure 8a. The interaction of heparin (sticks) with OPNs

positive core (blue) leads to a displacement of the N-terminal negatively charged region (red) either by direct repulsion through heparin or by loss of contacts to the positive region, explaining both the average increase in distances and the dynamic changes.

From isothermal titration calorimetry (ITC) measurements, ΔH and ΔS values of $-16.3 \text{ kCal mol}^{-1}$ and $-35 \text{ Cal mol}^{-1} \text{ K}^{-1}$, respectively, can be obtained for the heparin binding event of OPN assuming an average molecular weight of 17.5 kDa for heparin. These quite large ΔH and ΔS values nearly cancel each other at 293 K in the Gibbs energy ($\Delta G = \Delta H - T\Delta S$). Thus, although there is local rigidification in the heparin binding cleft (region around residue 155) the resulting conformational entropy penalty is reduced by a compensatory increase in conformational flexibility of the negatively charged region 90-120. The local (or segmental) unfolding and expansion of the OPN core segment thereby significantly contributes to the overall thermodynamic equilibrium balanced between counteracting contributions like solvation enthalpy, rotational and translational degrees of freedom and conformational entropy.^{48,49}

Conclusions

Three examples for weak and reversible protein interactions are presented in combination with magnetic resonance based experimental techniques to access these interactions. Relaxation dispersion (Car-Purcell-Meiboom-Gill, CPMG) experiments allow for determination of sparsely populated states in CspA-RNA complexation and MAX-MAX homodimerization. Since both kinetic data about complex formation (association and dissociation rates) as well as structural details of lowly populated (high-energy) conformational states are accessible by this technique, unique information about protein interactions becomes available. Further, the combination of paramagnetic relaxation enhancement (PRE) NMR and double electron-electron resonance (DEER) electron paramagnetic resonance (EPR) gives rise to the unprecedented and unexpected observation of cooperatively folded conformational states contained in the conformational ensemble of Osteopontin (OPN) and its changes in the weak OPN-heparin interaction. Our findings suggest that more elaborate

conceptual approaches are required for an adequate description of intrinsically disordered proteins and their conformational ensembles.

Proteins are characterized by significant structural plasticity and can undergo large structural rearrangements of the time-averaged conformational ensemble. It is therefore evident that classical structural biology approaches are only starting points for a comprehensive analysis of protein function. For intrinsically disordered proteins (IDPs) their flat energy landscapes allow for rapid exchange between different conformational isomers (substates). These often only sparsely populated conformational states, yet, lead to the formation of protein complexes that are essential for biological function. Detailed knowledge of the transiently formed intermediates along the reaction trajectory will be highly valuable. Further developments of new techniques but also the amendment of existing ones (as illustrated with the novel analysis technique of DEER data) will be necessary to provide a more complete picture of how proteins form biologically active molecular aggregates and perform the myriad of essential tasks and how this information can be exploited to manipulate their activities based on knowledge of the underlying chemical mechanisms.

Acknowledgements, Funding

We thank Prof. Dariush Hinderberger for his collaboration. D.K. acknowledges a Feodor Lynen scholarship by the Alexander von Humboldt-Foundation (Germany). This work was supported by the Austrian Science Foundation (FWF) (I844 to C.K. and R.K.).

References

- 1 Kumar, S., Ma, B. Y., Tsai, C. J., Sinha, N. & Nussinov, R. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Science* **9**, 10-19 (2000).
- 2 Loria, J. P., Berlow, R. B. & Watt, E. D. Characterization of enzyme motions by solution NMR relaxation dispersion. *Accounts of Chemical Research* **41**, 214-221, doi:Doi 10.1021/Ar700132n (2008).
- 3 Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789-796 (2009).
- 4 Tolman, J. R. & Ruan, K. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chemical Reviews* **106**, 1720-1736, doi:Doi 10.1021/Cr040429z (2006).
- 5 Bouvignies, G., Markwick, P., Bruschweiler, R. & Blackledge, M. Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings. *Journal of the American Chemical Society* **128**, 15100-15101, doi:Doi 10.1021/Ja066704b (2006).
- 6 Rule, G. S. & Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy*. (Springer, 2006).
- 7 Clore, G. M. & Iwahara, J. Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chemical Reviews* **109**, 4108-4139, doi:Doi 10.1021/Cr900033p (2009).
- 8 Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chemical Reviews* **104**, 3607-3622, doi:Doi 10.1021/Cr030403s (2004).
- 9 Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nature Structural Biology* **7**, 740-743, doi:Doi 10.1038/78963 (2000).
- 10 Kurzbach, D., Schwarz, T. C., Höfler, S., Hinderberger, D. & Konrat, R. *Angew. Chem. Int. Ed.* **accepted** (2014).
- 11 Platzer, G. *et al.* The Metastasis-Associated Extracellular Matrix Protein Osteopontin Forms Transient Structure in Ligand Interaction Sites. *Biochemistry* **50**, 6113-6124 (2011).
- 12 Tsai, C. J., Ma, B. Y., Sham, Y. Y., Kumar, S. & Nussinov, R. Structured disorder and conformational selection. *Proteins-Structure Function and Genetics* **44**, 418-427, doi:Doi 10.1002/Prot.1107 (2001).
- 13 Vallurupalli, P., Hansen, D. F. & Kay, L. E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *P Natl Acad Sci USA* **105**, 11766-11771, doi:DOI 10.1073/pnas.0804221105 (2008).

- 14 Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts Chem Res* **41**, 442-451, doi:Doi 10.1021/Ar700189y (2008).
- 15 Jen, J. Chemical Exchange and NMR T2 Relaxation - The Multisite Case. *J. Magn. Reson.* **30**, 111-128 (1978).
- 16 Tollinger, M., Skrynnikov, N. R., Mulder, F. A. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *Journal of the American Chemical Society* **123**, 11341-11352, doi:Doi 10.1021/Ja011300z (2001).
- 17 Mulder, F. A. A., Skrynnikov, N. R., Hon, B., Dahlquist, F. W. & Kay, L. E. Measurement of slow (μ s-ms) time scale dynamics in protein side chains by N-15 relaxation dispersion NMR spectroscopy: Application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J Am Chem Soc* **123**, 967-975, doi:Doi 10.1021/Ja003447g (2001).
- 18 Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1814**, 942-968, doi:Doi 10.1016/J.Bbapap.2010.10.012 (2011).
- 19 Tang, Y. *et al.* Fully automated high-quality NMR structure determination of small (2)H-enriched proteins. *J Struct Funct Genomics* **11**, 223-232, doi:10.1007/s10969-010-9095-6.
- 20 Xia, B., Ke, H. P., Jiang, W. & Inouye, M. The cold box stem-loop proximal to the 5'-end of the Escherichia coli cspA gene stabilizes its mRNA at low temperature. *Journal of Biological Chemistry* **277**, 6005-6011, doi:Doi 10.1074/Jbc.M109700200 (2002).
- 21 Newkirk, K. *et al.* Solution Nmr Structure of the Major Cold Shock Protein (Cspa) from Escherichia-Coli - Identification of a Binding Epitope for DNA. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5114-5118, doi:Doi 10.1073/Pnas.91.11.5114 (1994).
- 22 Shajani, Z. & Varani, G. NMR studies of dynamics in RNA and DNA by ¹³C relaxation. *Biopolymers* **86**, 348-359, doi:10.1002/bip.20650 (2007).
- 23 Thakur, C. S. & Dayie, T. K. Asymmetry of ¹³C labeled 3-pyruvate affords improved site specific labeling of RNA for NMR spectroscopy. *J Biomol NMR* **52**, 65-77, doi:10.1007/s10858-011-9582-5 (2012).
- 24 Lu, K., Miyazaki, Y. & Summers, M. F. Isotope labeling strategies for NMR studies of RNA. *J Biomol NMR* **46**, 113-125, doi:10.1007/s10858-009-9375-2 (2010).

- 25 Fauster, K., Kreutz, C. & Micura, R. 2'-SCF3 uridine-a powerful label for probing structure and function of RNA by ¹⁹F NMR spectroscopy. *Angew Chem Int Ed Engl* **51**, 13080-13084, doi:10.1002/anie.201207128 (2012).
- 26 Kreutz, C., Kahlig, H., Konrat, R. & Micura, R. A general approach for the identification of site-specific RNA binders by ¹⁹F NMR spectroscopy: proof of concept. *Angew Chem Int Ed Engl* **45**, 3450-3453, doi:10.1002/anie.200504174 (2006).
- 27 Kreutz, C., Kahlig, H., Konrat, R. & Micura, R. Ribose 2'-F labeling: a simple tool for the characterization of RNA secondary structure equilibria by ¹⁹F NMR spectroscopy. *J Am Chem Soc* **127**, 11558-11559, doi:10.1021/ja052844u (2005).
- 28 Wunderlich, C. H. *et al.* Synthesis of (6-(¹³C)pyrimidine nucleotides as spin-labels for RNA dynamics. *J Am Chem Soc* **134**, 7558-7569, doi:10.1021/ja302148g (2012).
- 29 Sauve, S., Tremblay, L. & Lavigne, P. The NMR solution structure of a mutant of the max b/HLH/LZ free of DNA: Insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *J Mol Biol* **342**, 813-832, doi:DOI 10.1016/j.jmb.2004.07.058 (2004).
- 30 Fieber, W. *et al.* Structure, function, and dynamics of the dimerization and DNA-binding domain of oncogenic transcription factor v-Myc. *J Mol Biol* **307**, 1395-1410, doi:DOI 10.1006/jmbi.2001.4537 (2001).
- 31 Kurzbach, D. *et al.* Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin. *Biochemistry* **52**, 5167-5175, doi:Doi 10.1021/Bi400502c (2013).
- 32 Drescher, M. EPR in Protein Science. *Top. Curr. Chem.* **321**, 91-120 (2012).
- 33 Kavalenka, A. *et al.* Conformational Analysis of the Partially Disordered Measles Virus N-TAIL-XD Complex by SDSL EPR Spectroscopy. *Biophys J* **98**, 1055-1064, doi:DOI 10.1016/j.bpj.2009.11.036 (2010).
- 34 Pirman, N. L., Milshteyn, E., Galiano, L., Hewlett, J. C. & Fanucci, G. E. Characterization of the disordered-to- α -helical transition of IA3 by SDSL-EPR spectroscopy. *Protein Sci* **20**, 150-159, doi:10.1002/pro.547 (2011).
- 35 Habchi, J. *et al.* in *Intrinsically Disordered Protein Analysis* Vol. 895 *Methods in Molecular Biology* (eds Vladimir N. Uversky & A. Keith Dunker) Ch. 21, 361-386 (Humana Press, 2012).
- 36 Kurzbach, D., Kattinig, D. R., Zhang, B. Z., Schluter, A. D. & Hinderberger, D. Loading and release capabilities of charged dendronized polymers revealed by EPR spectroscopy. *Chem Sci* **3**, 2550-2558, doi:Doi 10.1039/C2sc20435e (2012).

- 37 Junk, M. J. N., Spiess, H. W. & Hinderberger, D. The Distribution of Fatty Acids Reveals the Functional Structure of Human Serum Albumin. *Angew. Chem. Int. Ed.* **49**, 8755-8759 (2010).
- 38 Tsvetkov, Y. D., Milov, A. D. & Maryasov, A. G. Pulsed electron-electron double resonance (PELDOR) as EPR spectroscopy in nanometre range. *Russ. Chem. Rev.* **77**, 487-520 (2008).
- 39 Kurzbach, D., Kattinig, D. R., Pfaffenberger, N., Schärfl, W. & Hinderberger, D. Highly Defined, Colloid-Like Ionic Clusters in Solution. *ChemistryOpen* **1**, 211-214 (2012).
- 40 Jeschke, G., Panek, G., Godt, A., Bender, A. & Paulsen, H. *Appl. Magn. Reson.* **26**, 223-244 (2004).
- 41 Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *PCCP* **9**, 1895-1910 (2007).
- 42 Jeschke, G., Koch, A., Jonas, U. & Godt, A. Direct conversion of EPR dipolar time evolution data to distance distributions. *J Magn Reson* **155**, 72-82, doi:10.1006/jmre.2001.2498 (2002).
- 43 Jeschke, G., Koch, A., Jonas, U. & Godt, A. Direct conversion of EPR dipolar time evolution data to distance distributions. *J Magn Reson* **155**, 72-82, doi:DOI 10.1006/jmre.2001.2498 (2002).
- 44 Bellahcene, A., Castronovo, V., Ogbureke, K. U. E., Fisher, L. W. & Fedarko, N. S. Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer. *Nature Reviews Cancer* **8**, 212-226, doi:Doi 10.1038/Nrc2345 (2008).
- 45 Holde, K. E. v., Johnson, W. C. & Ho, P. S. *Principles of Physical Biochemistry*. (Prentice Hall, 1998).
- 46 Ponta, H., Sherman, L. & Herrlich, P. A. CD44: From adhesion molecules to signalling regulators. *Nature Reviews Molecular Cell Biology* **4**, 33-45, doi:Doi 10.1038/Nrm1004 (2003).
- 47 Clore, G. M., Tang, C. & Iwahara, J. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Current Opinion in Structural Biology* **17**, 603-616 (2007).
- 48 Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448**, 325-U323, doi:Doi 10.1038/Nature05959 (2007).
- 49 Uversky, V. N. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* **43**, 1090-1103 (2011).

Discussion

The historical bias towards stably folded proteins in structural biology is fading quickly as today genetic sequencing precedes all structural analysis. This has led to high research interest in the field of intrinsically disordered proteins (IDPs). The typical approach towards structure elucidation via X-ray crystallography is not feasible for these proteins. Additionally, their biophysical characterization often requires adaptation of existing methodology. In this thesis I present the efforts of our group regarding the elucidation of structural, dynamic and thermodynamic aspects in this highly relevant class of proteins.

In *Chapter 1*, we build on previous work carried out in this laboratory to demonstrate how the structural ensemble of OPN can be described using state of the art magnetic resonance methods. We are able to characterize lowly populated compact states and even demonstrate that some of them are of cooperatively folded nature.

In *Chapter 2*, we extend our approach to the analysis of the interaction of OPN with heparin. Here we can see the more compact ensembles (as demonstrated in *Chapter 1*) in action. Additionally, we demonstrate that the interaction of OPN with heparin leads to partial rigidification, while larger parts of the protein gain mobility upon interaction. Knowledge of the conformational ensemble of this IDP in its bound and unbound states aids us in our explanation of the observed binding behavior.

Chapter 3 demonstrates the utility of new magnetic resonance methodology in both, the detection and characterization of lowly populated states and the characterization of transient complexes. The utility of the CPMG pulse sequence is demonstrated for examples of protein-protein and protein-RNA interactions and we also demonstrate the possibility of studying the interaction as seen from either partner. The detection of weak interactions and lowly populated states not accessible by CPMG are demonstrated using the OPN-heparin interaction and the methodology developed for its elucidation is explained.

Chapter 1

As we have shown in our previous publication, Osteopontin can unequivocally be classified as an IDP.¹ In the same paper, we also detected a substantial amount of partial structuring within the ensemble of Osteopontin. **Chapter 1** now demonstrates our approach to the characterization of these more compacted states. To achieve a more detailed picture of the proteins properties, we combined methods of nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR) spectroscopy. While investigations in solution-state NMR can only detect the average of all contributing species of the conformational ensemble, pulse-sequences of EPR spectroscopy can be carried out on a freeze-quenched ensemble and, thus, report on the sum of all different states, rather than their time-average. Thus, these two techniques ideally complement each other in the case of IDP structure-elucidation.

The four single-mutants of Osteopontin available from previous studies (C54, C108, C188 and C247) were combined in such a way as to yield six different double mutants (C54-C108, C54-C188, C54-C247, C108-C188, C108-C247 and C188-C247). With these double mutants we could record double electron-electron resonance (DEER) spectra. This methodology measures the coupling of the unpaired electrons and can report on the distances between them. As the correlations result in damped cosine modulated signals, the interpretation of DEER data gained from a diverse ensemble of structures deviates from the one performed on data showing only relatively few distances. Theoretically a single distance between the two spin labels leads to a simple damped cosine-modulation, which can be detected and related to the distance. The distance distributions observed in IDPs, however, lead to the superposition of damped cosines whose summation results in signal decay. Although at first glance this simple decay might seem very uninformative, it has very interesting properties as depending on the composition of the distances present in the ensemble, the signal decays to different values at a given detection time. Importantly, as the measurement is carried out on a freeze-quenched ensemble we do not see a time-averaged signal of the ensemble, but a sum of the individual contributions of each state. Since small (inter-electron) distances contribute more strongly to the signal

decay than large distances, even low populations of compact states within the ensemble can be detected.

By applying this technique to OPN (and varying the concentration of the denaturant urea), we were able to demonstrate a sigmoidal behavior in the denaturation profile of Osteopontin. As sigmoidal behavior is a hallmark in cooperatively folded structures, we have, thus, demonstrated the presence of substructures within the observed ensemble of the IDP Osteopontin, which are cooperatively folded. It should be noted that a gradual and uniform increase in the distances in all populated structures present in the ensemble would not lead to the observed sigmoidal behavior. Interestingly the denaturation with 8M urea did not lead to a complete loss of compact structures within the ensemble of OPN. The addition of large amounts of salt was required in order to achieve complete loss of any short-distance correlations in the DEER data. Thus, we can conclude that the stabilization of compact (and surprisingly stable) states within OPN is achieved through both hydrophobic and electrostatic interactions.

Urea unfolding of OPN was additionally probed by paramagnetic relaxation enhancement measurements (PREs) for individual single mutants. In addition to the measurements at standard conditions, we also recorded PREs for mutant C188 in the presence of high concentrations of salt or urea. When we compare the three sets of PREs to each other, we can clearly see, that both urea and NaCl lead to a change in the conformational ensemble of OPN. The regions affected, however, are not the same in these two cases. While urea leads to an expansion of the region 110-160, NaCl mainly affects the region 70-120. In addition to the PRE measurements we also recorded the urea-gradient with ^{15}N - ^1H HSQC measurements. In these measurements we can again see the same regions being affected. Although a slight deviation from linear dependence to urea concentration can be observed, we have to refrain from interpreting this in terms of a cooperative denaturation behavior as the complex contributions to chemical shifts and the time averaged ensemble allow for other explanations for the observed sigmoidality.

Using a ^{13}C -methylation scheme for lysine residues we also probed the influence of urea on the chemical environment of side chains. The resulting data corroborate our previous findings. Taken together, our observations result in a model

Discussion

for Osteopontin that describes it as an ensemble of structures with a significant amount of compacted states. These compact structures are cooperatively folded and their compactness is highest in the core regions of OPN (100-180). Both hydrophobic interactions and electrostatic ones contribute to the formation of this core, with differing contributions as observed in Δ PREs. The latter observation can better be appreciated, when looking at the charge distribution along the primary sequence of Osteopontin.

Here, we have combined the intrinsically local probing of NMR techniques, with EPR measurements that allow for the observation of interactions over longer distances. Using both EPR/DEER data and NMR derived observations we were able to demonstrate the existence of cooperatively folded sub-states within the conformational ensemble of OPN. We also located the regions in OPN that contribute the strongest to these cooperatively folded sub-states and rationalized how they do so. Summarizing our findings we can say that:

- The conformational ensemble of OPN, although quickly interchanging and mainly consisting of ‘random-coil-like’ conformations, contains compact structures.
- Some of these compact structures are cooperatively folded
- The compact core of OPN is stabilized through electrostatic as well as hydrophobic interactions.

Chapter 2

In *Chapter 1*, we have demonstrated the applicability and advantages of combining EPR and NMR techniques to elucidate fundamental questions in IDPs. Our approach was centered on the solution state ensemble of the IDP Osteopontin and we have used denaturing conditions together with a combined magnetic resonance approach to elucidate specific features of its more compact substates. In **Chapter 2**, we apply these techniques to the interaction of Osteopontin with its natural ligand heparin. In order to generate an even more detailed picture of Osteopontin and its interaction with heparin, we additionally employed isothermal titration calorimetry

(ITC), nuclear spin relaxation measurements (R_1 , R_2 and hetNOE), diffusion ordered spectroscopy measurements (DOSY) and continuous wave electron paramagnetic resonance measurements (CW-EPR).

Many intrinsically disordered proteins fulfill functions within the dynamic interaction networks of the cell. Osteopontin is no exception and interacts with various binding partners.^{2,3,4,5} The biological relevance of the Osteopontin-heparin interaction stems from the fact that analogous interaction modes exist for the OPN-heparan sulfate and/or -hyaluronic acid complexes. These interactions are involved in the association of OPN with the CD44 receptor, which in turn are involved in cell signalling and adhesion.⁶

We collected PRE data sets in presence and absence of heparin for four cysteine single mutants mentioned previously (C54, C108, C188 and C247). When looking at the differential PRE effects one can clearly see an expansion of Osteopontin upon binding to heparin. Diffusion ordered spectroscopy (DOSY) measurements also confirmed the binding event between Osteopontin and heparin through an increase in the observed hydrodynamic radius. As more detailed information about the binding mode is, however, not accessible by DOSY measurements, we employed PRE measurements. It is important to note that PRE effects report on ensemble averaged data and the interconversions between the different substructures of the ensemble are very fast. As the PREs report on spatial proximity to a specific position within the primary sequence, the four mutants measured do not only tell us that a general expansion of Osteopontin takes place, but also that this expansion mainly happens in the region of residue positions 120-200 for the label at C108, while the label at position C188 reports the strongest effects between positions 70 and 120. When observing the transverse relaxation rate R_2 of ^{15}N and the heteronuclear Overhauser effect (hetNOE) between ^{15}N and ^1H in the backbone amides, these same regions display significant changes. While the region 150-160 displays decreased mobility and, thus, a rigidification of the protein, the central segment 90-130 shows an increase in mobility.

Continuous wave (CW) electron paramagnetic resonance measurements were carried out on the four single mutants as well. Again we saw an increase in mobility for the label in the region 90-150 (C108, $\tau_c=0.52\text{ns}$ / $\tau_{c,\text{heparin}}=0.43\text{ns}$) while the label

Discussion

close to position 160 (C188, $\tau_c=0.69\text{ns}$ / $\tau_{c,\text{heparin}}=0.87\text{ns}$) displays decreased mobility. The two labels C54 and C247 do not display significant changes in their correlation times upon heparin binding. Analysis of DEER data measured on Osteopontin double mutants was carried out using methodology established in *Chapter 1*. As in our system the effective modulation depth, Δ_{eff} , can be used as a reporter of interspin distances, we compared DEER measurements with and without heparin to gain additional information on the updated structural ensemble of OPN upon interaction. The six double mutants investigated all showed a decrease in Δ_{eff} , and thus an increase in distance, upon interaction with heparin. The strongest change in Δ_{eff} was observed for the double mutant 108-188, further corroborating the idea of an expansion of region 90-130 upon interaction. The specific charge distribution of Osteopontin, which was already mentioned in *Chapter 1*, can give us further indications as to why the observed changes occur. As heparin carries the highest negative charge density of any known biological macromolecule, an interaction with Osteopontin would be expected in the mainly positively charged region 145-165. The negatively charged region 75-125, however, also displays drastic changes. Remembering the conclusions drawn from *Chapter 1*, we know that the negatively and positively charged patches of Osteopontin interact with each other in the apo-state. If we consider the patches themselves to be stabilized by hydrophobic interactions, while their interaction with each other is predominantly electrostatic in character, we can expect the high amount of additional charge brought by an interaction with heparin to have dramatic effects. Indeed the positive patch of OPN interacts with the negatively charged heparin, thus rigidifying its compaction as seen in ΔR_2 , while the negatively charged patch loses its contact with the positive charges and is expelled from the heparin binding site as seen in ΔPREs . The loose attachment (and subsequent increased mobility) of the negatively charged regions in the complex is further supported by nuclear spin relaxation data. Together with the observation that Δ_{eff} decreases for the double mutant 54-108, we can further conclude that the negative patch itself becomes less compact upon interaction with heparin.

The relevance of electrostatics for the OPN-heparin interaction was further demonstrated with DOSY, PRE, ITC and chemical shift measurements under high salt conditions, all of which demonstrate loss of interaction.

Isothermal titration calorimetry measurements of Osteopontin with heparin were also carried out and yielded a dissociation constant of roughly 34 μM . Interestingly, this affinity is achieved with very high contributions from both ΔS and ΔH , which nearly cancel each other at ambient temperature. Additionally a positive differential heat capacity ΔC_p was observed when performing ITC measurements at various temperatures. Positive differential heat capacity changes are regularly found for protein unfolding measurements, while interactions that require folding upon binding are associated with negative ΔC_p . Increases in entropy of the protein chain and positive ΔC_p for the hydration of apolar groups confer the positive effect observed with protein unfolding.⁷ As in the interaction of Osteopontin and heparin the protein is already hydrated, displays many polar groups and the interaction is being driven by electrostatics, it is unlikely that we can see positive contributions from hydration. Thus, the observation of a positive ΔC_p is best explained by an increase in conformational flexibility, supporting our conclusion of increased overall flexibility upon interaction..

We have demonstrated that for the IDP Osteopontin segmental expansion and an increase in flexibility are observed upon interaction with heparin. This ‘unfolding upon binding’ adds to the vast repertoire of IDP interaction mechanisms and is most likely used to tune the contributions of entropy to binding events in various electrostatically governed IDP interactions. Summarizing this paper we would like to note that:

- Osteopontin displays a specific interaction with heparin.
- The conformational ensemble of OPN is considerably changed upon interaction.
- In contrast to a slight compaction of the binding site residues, other parts of OPN loose compaction and gain mobility upon interaction with heparin.
- These increases in mobility contribute (favorably) to the entropy of binding.
- We have demonstrated a novel ‘unfolding upon binding’ mechanism for this IDP.

Chapter 3

In **Chapter 3**, we describe the methods used in magnetic resonance spectroscopy to gain information on transiently formed protein complexes. The methods themselves are introduced briefly and research examples are presented for each of them. As IDPs are often involved in signalling and, thus, typically display high values for their dissociation constants,⁸ their interactions are very often transient in nature. This is not to say that only IDPs display transient interactions, they are also found for well-folded proteins. Both IDP and folded protein transient interactions play key roles in processes of molecular biology, such as post-translational modifications, cellular trafficking, transcription, translational regulation and cell signalling.

As transiently interacting protein complexes are not readily crystallized, information on their structures from x-ray diffraction is all but completely absent. NMR, on the other hand, can access these transiently formed complexes much more readily. The standard methodologies of chemical shift mapping, resonance intensity tracking and saturation techniques can be used to determine the interaction sites and also often allow for quantification of the dissociation constant. Efforts to perform detailed structural characterization of transient complexes by NMR are usually hampered by severe line broadening caused by the unfavorable exchange rate of these complexes. However, this complicity can be turned into a strength and source of information. Depending on the exchange rate, minor (high energy) states can have relatively drastic effects on the observed average. The transverse relaxation rate R_2 is particularly strongly affected by exchange on the high microsecond to mid millisecond timescale. In Car-Purcell-Meiboom-Gill (CPMG) measurements, this effect can be quenched depending on the repetition frequency of the applied CPMG pulse-train. By recording spectra with varying CPMG-frequency (ν_{CPMG}) we can fit the observed transverse relaxation rate ($R_{2,\text{eff}}$) as a function of ν_{CPMG} and, thus, the exchange rate between low energy (major) and the high energy (minor) states for each spin can be delineated as well as their relative quantities. Usually, a global fit for all residues reveals the exchange rate between the two states. As the conformational exchange shows variation on the residue level (for example through flipping of aromatic side

chains) a detailed analysis of the ν_{CPMG} dependence of $R_{2,\text{eff}}$ for each observed spin can lead to further insights in the structural and dynamic nature of the minor state. When recording these series of CPMG spectra at various magnetic field strengths, even the chemical shift of the minor state is accessible.

The applicability and achievable insights of the application of CPMG sequences to systems interchanging between major and high-energy states have already been demonstrated in several studies.^{9,10,11} In our publication we focus on its applicability to Protein-Protein and Protein-RNA interactions. The first application demonstrated is on the interaction of the *E.coli* cold-shock protein CspA with a fragment of its mRNA. The RNA chaperoning function of CspA is conveyed by its interaction with the 5'-untranslated region of its own mRNA (cold-box RNA, or CB RNA). For this study a partly complementary region to CB RNA, called anti-cold box RNA (ACB RNA) was used. The free state structure of this protein has been solved previously,¹² but the bound state was not. Accessing it through NMR proved unfeasible due to drastic signal loss. CPMG experiments using a ratio of 10:1 in CspA:ACB-RNA, however, could be recorded at various spectrometer frequencies. At this molar ratio a global fit determined the population of the bound state to about 1.6% of the conformational ensemble, with an exchange frequency of 1.4kHz. Residues close to the binding site had low signal to noise ratios and were not included in the global fit. Other residues (i.e. F12 and F18) however, showed clear CPMG profiles and as they are affected by large structural rearrangements upon binding, the detected frequencies can be correlated with the exchange rate (between free and bound).

The dynamics of the CspA interaction partners, the CB and ACB RNAs themselves, were also studied using CPMG pulse schemes. ^1H - ^{15}N correlation spectra of RNAs do not yield interpretable data, as the nitrogen bound protons in RNA bases undergo extremely fast exchange with water and are, thus, not observable in an aqueous environment. Therefore, ^{13}C based CPMG pulse schemata are usually employed to study the μs to ms dynamics of RNAs. Uniform labeling of RNA with ^{13}C would create a multitude of neighboring ^{13}C atoms and, therefore, lots of one-bond scalar couplings. As these couplings would lead to drastically more complex spectra as well as artifacts in the CPMG pulse sequences, more elaborate labeling schemes need to be used. In previous publications it was demonstrated how solid-phase synthesis can

Discussion

be used to achieve specific labeling patterns.^{13,14,15} It can be used to incorporate ^{19}F , which has the advantage of a much higher intrinsic NMR sensitivity than ^{13}C and a very broad chemical shift range. Using ^{19}F labels at specific positions, both the interaction and dynamics of anti coldbox RNA and coldbox RNA can be studied. We also carried out the more labor-intensive production of ^{13}C -modified pyrimidine phosphoramidites. Here no artificial chemical moiety is introduced, so we gain functionally unchanged RNA molecules. CPMG experiments with both, ^{19}F and ^{13}C labeled samples, were carried out and indicate a dynamic hotspot in the vicinity of the G•U wobble base pair. We have demonstrated the utility of CPMG experiments performed on ^{19}F and ^{13}C nuclei in addition to the already mentioned ^{15}N based ones. Using specific labeling schemes, we demonstrated the NMR accessibility of this RNA-protein interaction from the viewpoint of both interaction partners.

A third example to the applicability of CPMG pulse sequences in order to characterize transient complex formation is presented in form of the Myc-associated factor X (MAX) homodimerization. When interacting with its binding partner Myc the heterodimer Myc-Max forms and yields an active transcription factor involved in several severe diseases.¹⁶ The homodimerization of MAX can be influenced by pH and temperature, changing the monomer-dimer equilibrium. We have measured MAX in its monomeric form and used CPMG experiments to determine the fraction of homodimer in our conditions to be 0.75%. Despite this low percentage, we were able to accurately extract the ^{15}N chemical shift expected for the bound form, which can be measured, albeit under different conditions. As the chemical shift change is influenced by molecular rearrangements, CPMG allows us to obtain unique information on high-energy states even if their low population hinders access by other experimental techniques.

In the realm of IDPs the occurrence of transiently formed protein complexes is very common. As the interchange of the conformational ensemble of IDPs is typically faster than what can be observed with CPMG-type measurements, here we also present a set of techniques that is applicable to faster exchange. With this methodology we do not only gain information on interaction of IDPs, but we can also use it to elucidate specific features of the IDP's conformational ensemble.

In this article we have presented how CPMG methodology can be used to measure interesting features of transient protein complexes. The usage of these techniques was presented on several examples including protein-RNA, protein-protein and protein-carbohydrate complexes. Transiently formed intermediates are crucial for a vast amount of complexes formed in biology. The spectroscopic techniques presented here allow for the investigation of lowly populated states, be it transiently formed complexes or conformational intermediates, that can not be accessed by other, more conventional, methodology. Thus, we would like to emphasize that:

- CPMG measurements can access states populated as low as 0.75%.
- The calculation of the exchange rates between bound and free states is possible.
- Chemical shifts of amide nitrogens in the high energy state can be calculated and give insights into changes of the chemical environment of these amides upon interaction.
- CPMG measurements are possible with several heteronuclei and can be carried out on both protein and RNA samples.
- Information on μ s to ms dynamics is accessible with CPMG

Related Work

In the papers mentioned above, we have used mainly experimental techniques to derive information on the variability of IDPs and its relevance for binding events. Although we always look at experimental data to inform our understanding of IDPs, we have also used computational approaches in order to improve descriptions of IDPs.

Protonation-Dependent Conformational Variability of Intrinsically Disordered Proteins¹⁷

When using NMR to study proteins, a low pH in comparison to physiological conditions is usually beneficial to the signal to noise ratio as it slows down the amide proton exchange with water. This is especially true for IDPs as all of their amides are exposed to water. When looking at IDPs in a range of pH values, the spectra often demonstrate considerable non-linear changes. Interestingly higher aggregation

Discussion

propensities as well as increased α -helicity have already been observed in IDPs at low pH values.¹⁸ In addition, their net charge is an important factor contributing to overall compaction and hydrodynamic radii.¹⁹ We employed the meta-structure²⁰ approach to identify expected changes of IDPs in regard to the pH value. As the calculation of meta-structure values does only rely on the primary sequence, a relatively simple computational approach for simulating low pH on a large set of sequences could be employed. In order to mimic IDPs in low pH conditions, amino acids that are protonated at low pH values (Glutamate and Aspartate) were simply changed to amino acids which are chemically comparable but do not carry a charge at neutral pH (Glutamine and Asparagine). In contrast to the protonated forms of glutamate and aspartate, these electro-neutral forms are present in the set of structures used to generate the meta-structure pairwise distribution function $p(\theta, A, B, l_{AB})$.

The large-scale comparison of intrinsically disordered regions revealed a tendency towards the formation of α -helices as well as an overall compaction within previously disordered regions upon pH reduction. Although these calculations assume a pH level far lower than commonly encountered in live cells, the tight regulation of pH,²¹ as well as the drastic differences for local pH levels encountered within cells,²² indicate that the results obtained can have physiologically relevant consequences. Two additional factors contribute to this relevance: Firstly a complete protonation of the side chain is most likely not needed to change the conformational ensemble of the protein, as partial protonation can already reduce the electrostatic repulsion between acidic side chains. Secondly the protonation does not have to be carried out via reduction of pH, it can also be achieved via electrostatic interactions where the negatively charged side chain simply acts as partial electron donor and proton acceptor. In this way an interaction with charged species leads to the preferential formation of an α -helix without the necessity of a drastic change in pH.

Examples for this mechanism are the α -helix formation of α -Synuclein upon interaction with micelles²³ as well as in low pH conditions¹⁸ and the stabilization of partially pre-formed α -helices within the unstructured protein Tcf4 upon interaction with beta-catenin.^{24,25} In addition, we have demonstrated predictive capacity of this approach as shown with the increased α -helicity of BASP1 upon pH reduction. Thus, in addition to gaining information on compact states of specific IDPs and

demonstrating methods to investigate their interactions in the three publications that make up this thesis, we have also shown our capability in assessing the generality of mechanisms observed in select IDPs.¹⁷

Structure and Regulatory Interactions of the Cytoplasmic Terminal Domains of Serotonin Transporter²⁶

Examples discussed in this thesis generally demonstrate our capability to investigate intrinsically disordered proteins on a per-residue level after assignment of resonances has been achieved. For assignment one typically requires a large amount of highly concentrated recombinantly expressed ¹⁵N and ¹³C labeled protein. With the assignment at hand, we can give detailed insights into the conformational complexities of these proteins. Even in the absence of artificially introduced NMR-active nuclei, however, NMR can contribute to a detailed picture of proteins. In the case of the human serotonin transporter (SERT), for example, the structure of the N-terminal domain (82 amino acids) was investigated. Although the structure of the transmembrane region of SERT is not known, homology modeling was possible and provides a relatively reliable estimate for this region. The terminal regions are exposed to the aqueous environments of the cytosol and synaptic cleft and computational methods only predict low amounts of secondary structure for them.

In a collaborative study our group contributed ¹H NMR data sets as well as CD measurements to assess the structure of the 82 N-terminal residues of SERT. We demonstrated a lack of stable tertiary structure using simple proton spectra to exclude beta-sheet contributions and a stable core as seen in a lack of amide proton dispersion. A stable α -helix could be excluded using circular dichroism (CD) measurements, which are particularly well suited to detect α -helices as they lead to minima in ellipticity at 208 and 220 nm. Despite the lack of stable structure, diffusion ordered spectroscopy (DOSY) measurements allowed us to determine the radius of hydration to roughly 24.5 Å which is smaller than expected for a pure random coil like protein (29.6 Å), but larger than for a globular protein (17.8 Å) of the same primary sequence length.²⁷ We also used the meta-structure²⁰ approach to determine secondary structure propensities as well as expected compactness values. Together with bioinformatic predictors, mutagenesis, biochemical measurements and FRET microscopy, these data were used in *de novo* and homology modeling runs to generate models of full-length

SERT that are consistent with the above mentioned as well as previously obtained experimental data. The models also predict movements in the conformational cycle that conform to the ‘rocking bundle’ hypothesis. These movements are consistent with obtained spectroscopic data as well as with structures of homologues.²⁶

A De-Novo Disease-Modifying Compound that Targets α -Synuclein Improves Deficits in Models of PD/DLP

As alpha Synuclein (α -Syn) is involved in ‘Dementia with Lewy Bodies’, ‘Multiple System Atrophy’ and ‘Parkinson’s Disease’, which are all sicknesses with large impacts on today’s society, research interest in this protein has been substantial for the last decades.^{28,29} As it is also an intrinsically unstructured protein, several NMR based studies exploring its conformational ensemble in its free form as well as its more structured ensemble when membrane bound have been carried out.³⁰ In collaboration with our industry partner *Neuropore Therapies Inc.* and the group of Eliezer Masliah, Department of Neurology, UCSD, we have tried to elucidate the effects of the *de novo*-developed compound NPT100-18A on both of these states. To mimic membranes we employed liposomes generated from 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG). α -Syn binding to these liposomes leads to signal loss, caused by the large diameter and, thus, slow tumbling of the liposomes employed. Interestingly, the signal intensities plotted along the primary sequence of α -Syn display a distinct binding profile, which corresponds to two separate binding modes of α -Syn termed SL1 and SL2 in previous publications.³¹ We were able to demonstrate that the developed compound does not display measurable interactions with α -Syn in its free form, but it does displace the protein from the liposome. Additionally we can demonstrate that the binding mode primarily affected (SL2) is the one involving a larger region of α -Syn. Mutational studies demonstrated the importance of residue E83 for the interaction of the compound with liposome bound α -Syn. The compound was developed on the basis of computational models and aimed to interrupt a speculative dimer formation of α -Syn in the membrane. The region it was developed for is only membrane bound in binding mode SL2. In addition a change of E83 to K was also computationally predicted to lower the affinity of the compound to α -Syn. Our NMR based analysis of the compound has helped to substantiate the proposed mechanism of action of this compound which has been shown by our collaborators to ameliorate motor deficits and reduce synaptic accumulation of α -Syn in mouse models.

Map1B Light Chain and its Interaction with Microtubules

Another system we have studied employing methodology developed for α -Synuclein and Osteopontin is the interaction of the microtubule-associated protein MAP1B with microtubules. MAP1B is a neuronal protein that is involved in the maturation of synapses and in network formation in murine brain development.³² As it is also implicated in human disease, its function in general and its mode of interaction with microtubules in particular are of great interest. As a start into our investigations we have already managed to assign the backbone and many side-chain resonances of MAP1B in collaboration with the group of W. Kozminski.³³ The determined increased α -helical propensity at the N-terminus of the protein was mirrored in the meta-structure analysis of the protein. In this study we will use paramagnetic relaxation enhancements to describe the ensemble of free state structures of the protein. As we have already gathered first interaction data with microtubules, we are confident we can extend this analysis to the bound form as well. In addition to the analysis of relative heights as performed for α -Syn, we will also try to use the recently developed CEST/DEST^{34,35} approaches to determine a more detailed interaction of MAP1B with both: polymerized microtubules and tubulin dimers.

Concluding Remarks

Historically, the catalytic activity of enzymes and their suitability for X-ray diffraction studies made them the prime interest of structural biology. Both these traits are predominantly observed in well-folded proteins, leading them to dominate the field of structural biology. Today, however, the possibility of genome-wide sequencing and new techniques, based primarily on genetic manipulations, have made all coding sequences in the genome available to molecular biology studies. In this unrestricted set of sequences many proteins display amino acid compositions indicative of a lack of tertiary structure. These previously ignored intrinsically disordered proteins make up a large fraction of coding sequences in higher organisms and their structural characterization has only recently started.

Discussion

Indeed, many proteins do display a stably folded conformation and the description and models used so far have served us very well in describing them. This is especially true for detailed descriptions of active sites and how they can catalyze chemical reactions. The idea that a well-defined structure is required for protein function and that these two properties are closely linked has allowed for great progress in structural biology and many future achievements will still build on this paradigm. All proteins, however, display varying amounts of motion and many possess so-called intrinsically unstructured regions or are entirely devoid of stable tertiary structure. This, however, does not mean that these regions can simply be written off as 'unstructured' and therefore 'without function'. They have been shown to carry out essential tasks in the biological environment³⁶ and are often involved in signalling cascades,⁸ which makes them both prone to disease mutation and interesting targets for drug development. If we understand that it is our idea of structure and the measurement techniques we use, that made these proteins 'unstructured', rather their inherent conformations, we can start to develop new ways to describe their behavior. When we go away from the dichotomic (black and white) view about protein structure, we can appreciate the conformational complexity of these proteins and devise new ways to measure and describe them.

This is what we set out to do in the papers comprising this thesis and are still doing now. With a detailed description of many properties of the conformational ensemble of Osteopontin and the influence denaturing reagents have on it, we have demonstrated that this protein, which is commonly termed 'unstructured', does contain cooperatively folded structures within its ensemble. To achieve this, we have combined electron paramagnetic resonance (EPR) and nuclear magnetic resonance (NMR). In the case of EPR we were even able to show that commonly recorded double electron electron resonance (DEER) measurements, which were to this date only used for stably folded proteins, yield data that can be very useful in the realm of intrinsically disordered proteins (IDPs).

In addition to an improved description of the conformational ensemble of the free state of this protein, we have used similar techniques to characterize the changes that occur upon ligand binding. To our surprise we observed an expansion, which is likely used to compensate for the high entropy loss usually incurred by an IDP when it

encounters a binding partner. Although the conformational ensemble is expected to vary greatly between various proteins commonly classified as IDPs, the detailed description of this IDP should be helpful in improving our understanding of this class of proteins generally. The feature of cooperatively folded states within the ensemble, for example, is most likely not limited to Osteopontin and its demonstration here should help to update the way we view IDPs in general. Similarly, the observed expansion upon binding adds another mode of interaction to the arsenal of IDPs which in addition to ‘conformational selection’ already includes ‘fuzzy complexes’³⁷ and ‘folding upon binding’.³⁸ And the observed compaction of the binding site in at least some parts of the conformational ensemble strongly suggests that for IDPs at least partial ‘conformational selection’ is likely present even in interactions mostly following the proposed ‘folding upon binding’ mechanism.

We have demonstrated various techniques and how we used them to elucidate the conformational ensemble of IDPs. As these techniques are fine-tuned for the detection of lowly populated states and changes in them, they can also be used to study transient interactions. These interactions play crucial roles in biology,^{39,40,41} but they do pose problems for many conventional experimental techniques. In addition to the techniques used for IDPs we have also demonstrated the use of Car-Purcell-Meiboom-Gill (CPMG) measurements, which are ideally suited to detect fluctuations occurring on a μ s to ms timescale. Using this technique we have demonstrated how we can access both the exchange between states as well as gather detailed information on the excited states themselves.

In my thesis I have provided a detailed description of the structural ensemble and the dynamics of the IDP Osteopontin in its free and heparin-bound states. Additional thermodynamic data on the interaction and a rationale for the observed changes are presented. I also explain the methodologies developed on this model system as well as the CPMG pulse sequence using several examples. Thus, my efforts in contributing to the broad field of IDPs and transient interactions in general and the OPN-heparin interaction in particular, are presented in this thesis.

References

- 1 Platzter, G. *et al.* The metastasis-associated extracellular matrix protein osteopontin forms transient structure in ligand interaction sites. *Biochemistry* **50**, 6113-6124, doi:10.1021/bi200291e (2011).
- 2 Poggio, P. *et al.* Osteopontin-CD44v6 interaction mediates calcium deposition via phospho-Akt in valve interstitial cells from patients with noncalcified aortic valve sclerosis. *Arterioscler Thromb Vasc Biol* **34**, 2086-2094, doi:10.1161/ATVBAHA.113.303017 (2014).
- 3 Banerjee, A., Lee, J. H. & Ramaiah, S. K. Interaction of osteopontin with neutrophil alpha(4)beta(1) and alpha(9)beta(1) integrins in a rodent model of alcoholic liver disease. *Toxicol Appl Pharmacol* **233**, 238-246, doi:10.1016/j.taap.2008.08.008 (2008).
- 4 Marcinkiewicz, C. *et al.* Inhibitory effects of MLDG-containing heterodimeric disintegrins reveal distinct structural requirements for interaction of the integrin alpha 9beta 1 with VCAM-1, tenascin-C, and osteopontin. *J Biol Chem* **275**, 31930-31937, doi:10.1074/jbc.M003209200 (2000).
- 5 Barry, S. T., Ludbrook, S. B., Murrison, E. & Horgan, C. M. A regulated interaction between alpha5beta1 integrin and osteopontin. *Biochem Biophys Res Commun* **267**, 764-769, doi:10.1006/bbrc.1999.2032 (2000).
- 6 Ponta, H., Sherman, L. & Herrlich, P. A. CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol* **4**, 33-45, doi:10.1038/nrm1004 (2003).
- 7 Prabhu, N. V. & Sharp, K. A. Heat capacity in proteins. *Annu Rev Phys Chem* **56**, 521-548, doi:10.1146/annurev.physchem.56.092503.141202 (2005).
- 8 Tantos, A., Han, K. H. & Tompa, P. Intrinsic disorder in cell signaling and gene transcription. *Mol Cell Endocrinol* **348**, 457-465, doi:10.1016/j.mce.2011.07.015 (2012).
- 9 Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc Chem Res* **41**, 442-451, doi:10.1021/ar700189y (2008).
- 10 Vallurupalli, P., Hansen, D. F. & Kay, L. E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci U S A* **105**, 11766-11771, doi:10.1073/pnas.0804221105 (2008).
- 11 Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**, 942-968, doi:10.1016/j.bbapap.2010.10.012 (2011).
- 12 Tang, Y. *et al.* Fully automated high-quality NMR structure determination of small (2)H-enriched proteins. *J Struct Funct Genomics* **11**, 223-232, doi:10.1007/s10969-010-9095-6 (2010).
- 13 Kreutz, C., Kahlig, H., Konrat, R. & Micura, R. Ribose 2'-F labeling: a simple tool for the characterization of RNA secondary structure equilibria by 19F NMR spectroscopy. *J Am Chem Soc* **127**, 11558-11559, doi:10.1021/ja052844u (2005).

- 14 Kreutz, C., Kahlig, H., Konrat, R. & Micura, R. A general approach for the identification of site-specific RNA binders by ¹⁹F NMR spectroscopy: proof of concept. *Angew Chem Int Ed Engl* **45**, 3450-3453, doi:10.1002/anie.200504174 (2006).
- 15 Fauster, K., Kreutz, C. & Micura, R. 2'-SCF3 uridine-a powerful label for probing structure and function of RNA by ¹⁹F NMR spectroscopy. *Angew Chem Int Ed Engl* **51**, 13080-13084, doi:10.1002/anie.201207128 (2012).
- 16 Sauve, S., Tremblay, L. & Lavigne, P. The NMR solution structure of a mutant of the Max b/HLH/LZ free of DNA: insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *J Mol Biol* **342**, 813-832, doi:10.1016/j.jmb.2004.07.058 (2004).
- 17 Geist, L. *et al.* Protonation-dependent conformational variability of intrinsically disordered proteins. *Protein Sci* **22**, 1196-1205, doi:10.1002/pro.2304 (2013).
- 18 Cho, M. K. *et al.* Structural characterization of alpha-synuclein in an aggregation prone state. *Protein Sci* **18**, 1840-1846, doi:10.1002/pro.194 (2009).
- 19 Marsh, J. A. & Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys J* **98**, 2383-2390, doi:10.1016/j.bpj.2010.02.006 (2010).
- 20 Konrat, R. The protein meta-structure: a novel concept for chemical and molecular biology. *Cell Mol Life Sci* **66**, 3625-3639, doi:10.1007/s00018-009-0117-0 (2009).
- 21 Boron, W. F. Regulation of intracellular pH. *Adv Physiol Educ* **28**, 160-179, doi:10.1152/advan.00045.2004 (2004).
- 22 Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G. & Tsien, R. Y. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc Natl Acad Sci U S A* **95**, 6803-6808 (1998).
- 23 Rao, J. N., Jao, C. C., Hegde, B. G., Langen, R. & Ulmer, T. S. A combinatorial NMR and EPR approach for evaluating the structural ensemble of partially folded proteins. *J Am Chem Soc* **132**, 8657-8668, doi:10.1021/ja100646t (2010).
- 24 Poy, F., Lepourcelet, M., Shivdasani, R. A. & Eck, M. J. Structure of a human Tcf4-beta-catenin complex. *Nat Struct Biol* **8**, 1053-1057, doi:10.1038/nsb720 (2001).
- 25 Graham, T. A., Ferkey, D. M., Mao, F., Kimelman, D. & Xu, W. Tcf4 can specifically recognize beta-catenin using alternative conformations. *Nat Struct Biol* **8**, 1048-1052, doi:10.1038/nsb718 (2001).
- 26 Fenollar-Ferrer, C. *et al.* Structure and regulatory interactions of the cytoplasmic terminal domains of serotonin transporter. *Biochemistry* **53**, 5444-5460, doi:10.1021/bi500637f (2014).
- 27 Wilkins, D. K. *et al.* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424-16431 (1999).

- 28 Brookes, A. J. & St Clair, D. Synuclein proteins and Alzheimer's disease. *Trends Neurosci* **17**, 404-405 (1994).
- 29 Binolfi, A., Theillet, F. X. & Selenko, P. Bacterial in-cell NMR of human alpha-synuclein: a disordered monomer by nature? *Biochem Soc Trans* **40**, 950-954, doi:10.1042/BST20120096 (2012).
- 30 Eliezer, D., Kutluay, E., Bussell, R., Jr. & Browne, G. Conformational properties of alpha-synuclein in its free and lipid-associated states. *J Mol Biol* **307**, 1061-1073, doi:10.1006/jmbi.2001.4538 (2001).
- 31 Bodner, C. R., Maltsev, A. S., Dobson, C. M. & Bax, A. Differential phospholipid binding of alpha-synuclein variants implicated in Parkinson's disease revealed by solution NMR spectroscopy. *Biochemistry* **49**, 862-871, doi:10.1021/bi901723p (2010).
- 32 Meixner, A. *et al.* MAP1B is required for axon guidance and is involved in the development of the central and peripheral nervous system. *J Cell Biol* **151**, 1169-1178 (2000).
- 33 Orban-Nemeth, Z. *et al.* Backbone and partial side chain assignment of the microtubule binding domain of the MAP1B light chain. *Biomol NMR Assign* **8**, 123-127, doi:10.1007/s12104-013-9466-6 (2014).
- 34 Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* **134**, 8148-8161, doi:10.1021/ja3001419 (2012).
- 35 Fawzi, N. L., Ying, J., Ghirlando, R., Torchia, D. A. & Clore, G. M. Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. *Nature* **480**, 268-272, doi:10.1038/nature10577 (2011).
- 36 Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, 553-584, doi:10.1146/annurev-biochem-072711-164947 (2014).
- 37 Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, 2-8, doi:10.1016/j.tibs.2007.10.003 (2008).
- 38 Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60 (2002).
- 39 Kumar, S., Ma, B., Tsai, C. J., Sinha, N. & Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* **9**, 10-19, doi:10.1110/ps.9.1.10 (2000).
- 40 Loria, J. P., Berlow, R. B. & Watt, E. D. Characterization of enzyme motions by solution NMR relaxation dispersion. *Acc Chem Res* **41**, 214-221, doi:10.1021/ar700132n (2008).
- 41 Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789-796, doi:10.1038/nchembio.232 (2009).

Acknowledgements

Acknowledgements

As it is the custom in any academic paper to mention everybody that contributed to this publication, I would like to thank all the people who have helped me in one way or another in producing this thesis. Indeed, none of the research I demonstrated in this cumulative dissertation as well as the work itself would have been possible without the generous help of my co-workers, friends and family.

I would like to thank my PhD-supervisor **Robert Konrat** for the many opportunities he gave me or helped me take advantage of. I am certain that there are not many professors who would allow a student who has just ruined a lab centrifuge to start a Masters thesis. He helped steer all of my research efforts in fruitful directions and also showed me how important it can sometimes be to accept that a protein will sometimes not do as you please, in order to be able to focus on new research questions.

I am grateful for all the scientific feedback from my PhD committee members **Gang Dong**, **Kristina Djinovic-Carugo** and **Bojan Zagrovic** that helped steer me towards the completion of my PhD. The administrative assistance of **Karin Pfeiffer** and **Ulrike Seifert** and, especially in the final phases of my PhD, **Gerlinde Aschauer** was absolutely invaluable for me in order to successfully complete my degree.

It was a wonderful experience to work with a research group that is very dedicated to their scientific efforts but also supportive in personal matters. I would like to thank **Nicolas Coudeville** for being my ‘training wheels’ in regards to NMR during my masters and for continuing to give me valuable inputs all through my PhD. I want to thank **Leonhard Geist**, **Gerald Platzer**, **Tomas Sara** and **Gönül Kizilsavas** for their continued support in scientific matters and the hours we spent trying to figure out particulars of NMR-theory and how to get high quality samples. I want to thank **Georg Kontaxis** for all his help regarding everything related to pulse-sequences and **Dennis Kurzbach** for his input in everything regarding EPR as well as for the ceaseless generation of nicknames for all lab members and PIs. I want to thank **Andrea Flamm** for lots of help in regard to liposome production. All lab members were always helpful when questioned and helped me with advice and actual protein samples. I want to give a special thank you to **Karin Ledolter** for being the reason that our lab always felt like a great workplace and for all the help in protein production and purification she has given me. Without her much of the work I did at the lab would not have been possible and she was essential to the research success I enjoyed so far.

I would like to thank my father **Wolfgang** for his continued support throughout the long years of my education and for the many discussions that have trained me well for academia, where every notion and result is questioned. I want to thank my mother **Brix** for giving me the foundation of an inquisitive mind that made me want to pursue a postgradual education and for all the emotional support I received throughout it. My siblings **Christoph**, **Veronika**, and **Sophie** I want to thank for understanding the problems I faced and for all the times we spent just chatting and forgetting about work. I also want to thank my grandparents **Ludwig** and **Hedwig** for the many Sunday lunches I thoroughly enjoyed and used for lively discussions as well as relaxation. My grandfather **Herbert** I want to thank for always being understanding and for hosting wonderful Christmas days with the extended family.

A great thank you also goes to my friends **Thomas**, **Petra**, **Sarah**, **Christine** and **Claudia**. I fondly remember the many hours spent talking, watching movies or playing games. All of them were a great source of support by helping me relax and listening to my problems.

Most of all I would like to thank my girlfriend **Christina** who has had the doubtful pleasure of listening to all my problems and still supported me in all my efforts. Who put up with my forgetful and talkative nature and balanced my world in every way imaginable. Who is the greatest source of positive energy in my life and who will stick by me come what may, I cherished every minute I spent with you in the last seven years and I will cherish all those that come.

Curriculum Vitae

Personal Information

First Name: **Thomas Christian**
Last Name: **Schwarz**
Academic Titles: BSc, MSc
Nationality: Austria

Education

Mar 2011 – Apr 2015 **Doctoral Thesis**, Dept. of Structural and Computational Biology, MFPL, University of Vienna, Austria

Mar 2011 – Mar 2012 Received scholarship '**Forschungsstipendium 2011**' awarded by the University of Vienna, Austria

Mar 2010 – Feb 2011 **Master Thesis**, Dept of Structural and Computational Biology, MFPL, University of Vienna, Austria. Received title Master of Science

Mar 2010 – Feb 2011 **Master Studies Molecular Biology**, University of Vienna, Austria

Nov 2008 – Dec 2009 **Bachelor Studies Biology**, University of Vienna, Austria. Received title Bachelor of Science

Feb 2009 – June 2009 **Joint Study** - Exchange Student, Departamento de Biología Molecular de Plantas, IBT, Universidad Nacional Autónoma de México, Mexico

Jan 2008 – Jun 2008 **Erasmus** - Exchange Student, Department of Chemistry, Warwick University, England

Oct 2004 – Feb 2010 **Diploma Studies Molecular Biology**, University of Vienna, Austria

Laboratory Experience

Since Mar 2011 **Doctoral Thesis** under supervision of Prof. Robert Konrat (MFPL) NMR

Summer term 2010 **Master Thesis** under supervision of Prof. Robert Konrat (MFPL) NMR

Winter term 2009/10 **Student Project** with Prof. Gang Dong (MFPL) Crystallography

Summer term 2009 **Joint Study** - Student Project with Prof. Omar Pantoja (UNAM) Plant Molecular Biology

Summer term 2008 **Erasmus** - Student Project with Prof. Hendrik Schaefer (Warwick) Microbiology

Summer 2007 **Student Project** with Prof. Kristina Djinovic-Carugo (MFPL) Crystallography

Language Skills

German: mother tongue
English: fluent
Spanish: intermediate

Conferences and Meetings

Mar 2015 Oral Presentation – MMCE, Krynica-Zdroj, Poland
 Jun 2014 Poster Presentation – Euromar, Zurich, Switzerland
 Mar 2013 Poster Presentation – MMCE, Semmering, Austria
 Aug 2012 Poster Presentation – ICMRBS, Lyon, France

Publications

Sara, T., Schwarz, T.C., Kurzbach, D., Wunderlich, C.H., Kreutz, C., Konrat, R. Magnetic Resonance Access to Transiently Formed Protein Complexes. *Chemistry Open* **3**, 115-123 (2014)

Kurzbach, D.,* Schwarz, T.C.,* Platzer, G., Hoefler, S., Hinderberger, D., Konrat, R. Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex. *Angewandte Chemie International Edition* **53**, 3840-3843 (2014)

*both authors contributed equally

Fenollar-Ferrer, C., Stockner, T., Schwarz, T.C., Pal, A., Gotovina, J., Hofmaier, T., Jayaraman, K., Adhikary, S., Mehdipour, A.R., Kudlacek, O., Rudnick, G., Tavoulari, S., Singh, S.K., Konrat, R., Sitte, H.H., Forrest, L.R. On the Structure and Regulatory Interactions of the Cytoplasmic Terminal Domains of Serotonin Transporter. *Biochemistry* **53**, 5444-60 (2014)

Platzer, G.,* Kurzbach, D.,* Schwarz, T.C., Henen, M.A., Konrat, R., Hinderberger, D. Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin. *Biochemistry* **52**, 5167-5175 (2013)

*both authors contributed equally

Geist, L., Henen, M.A., Haiderer, S., Schwarz, T.C., Kurzbach, D., Zawadzka-Kazimierczuk, A., Saxena, S., Zerko, S., Kozminski, W., Hinderberger, D., Konrat, R. Protonation-dependent Conformational Variability of Intrinsically Disordered Proteins. *Protein Science* **22**, 1196-19205 (2013)