



DISSERTATION

Titel der Dissertation

„Osteopontin and the molecular principles that
orchestrate its rheomorphic ensemble“

Verfasser

Gerald Platzer

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2013

Studienkennzahl lt. Studienblatt:

A 094 490

Dissertationsgebiet lt. Studienblatt:

Molekulare Biologie

Betreuerin / Betreuer:

Prof. Dr. Robert Konrat

This cumulative dissertation is based on the following publications:

The Metastasis-Associated Extracellular Matrix Protein Osteopontin Forms Transient Structure in Ligand Interaction Sites

Platzer G, Schedlbauer A, Chemelli A, Ozdowy P, Coudevylle N, Auer R, Kontaxis G, Hartl M, Miles AJ, Wallace BA, Glatter O, Bister K, Konrat R.

Biochemistry **2012** 50(27), 6113-6124

Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin

Platzer G,* Kurzbach D,* Schwarz TC, Henen MA, Konrat R, Hinderberger D.

Biochemistry **2013** 52(31), 5167-5175

*both authors contributed equally

Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex

Kurzbach D,* Schwarz TC,* Platzer G, Höfler S, Hinderberger D, Konrat R.

Angewandte Chemie **2013**, under revision

*both authors contributed equally

Abstract

Proteins perform a vast array of functions within a living organism ranging from the catalysis of metabolic reactions and DNA synthesis to cellular transport, signaling and many others. While many proteins rely on a well-defined tertiary structure to fulfill their physiological roles, a large number of proteins in higher organisms are found to contain unstructured regions or in extreme cases display a total lack of any tertiary structure in solution. These so-called intrinsically disordered proteins (IDPs) have recently attracted a lot of attention in the scientific community given their interesting biophysical properties and their involvement in a multitude of physiological and pathological processes.

In this work we investigate the dynamic and structural properties of the IDP Osteopontin (OPN). OPN is a highly negatively charged polypeptide and is a target of several post-translational modifications (PTMs) that govern its function *in vivo*. It is expressed in a variety of cell-types and depending on its localization, is involved in variety of cellular events including biomineralization, immune response, extracellular matrix (ECM)-degradation and cell signaling. Additionally, the protein is overexpressed in malignant tissues where it exerts its tumorigenic function by interacting with cell-surface receptors like Integrins and CD44, therefore rendering it a highly interesting research target.

Since X-ray crystallography, as the classical approach for protein-structure determination is not applicable in the case of IDPs, we use a methodologically integrated approach of NMR, EPR, SAXS and CD-spectroscopy to gain insights into the highly flexible nature of OPN. Our findings show that OPN displays a highly diverse conformational ensemble comprising cooperative folding and unfolding events between extended and compact structures in solution. Finally, we address the implications of disorder on the interaction with Osteopontin's naturally occurring ligand heparin.

Zusammenfassung

Proteine sind an einer Vielzahl von Prozessen in lebenden Organismen beteiligt. Diese reichen von der Umsetzung von Metaboliten, Transformation, Replikation und Reparatur von DNS, über zellulären Transport bis hin zu Signaltransduktionsprozessen. Obwohl viele Proteine auf eine wohldefinierte dreidimensionale Struktur angewiesen sind um ihre physiologische Rolle auszuüben, finden sich in höheren Lebewesen viele Proteine die teilweise unstrukturierte Bereiche enthalten bis hin zu komplett unstrukturierten Proteinen. Diese sogenannten Intrinsisch Ungefalteten Proteine (IUPs) haben in den letzten Jahren viel Interesse geweckt, besonders angesichts ihrer Beteiligung an einer großen Anzahl an physiologischen und pathophysiologischen Prozessen.

In dieser Arbeit untersuchen wir die dynamischen und strukturellen Eigenschaften des IUPs Osteopontin (OPN). OPN ist ein hoch negativ geladenes Polypeptid und Ziel einer Vielzahl von post-translationalen Modifikationen, die ausschlaggebend für seine in vivo Funktion sind. OPN wird von vielen verschiedenen Zellen exprimiert und ist, je nach zellulärer Lokalisation an unterschiedlichen Prozessen beteiligt, unter anderem Biomineralisierung, Immunantwort, Abbau von Komponenten der extrazellulären Matrix und Signaltransduktionsprozessen. Zusätzlich wird OPN in vielen Tumoren überexprimiert wo es durch Interaktion mit Zellrezeptoren wie Integrinen und CD44 in tumorfördernden Prozessen wechselwirkt.

Da Röntgendiffraktometrie, die klassische Methode zur Strukturbestimmung von Proteinen im Falle von IUPs nicht anwendbar ist, verwenden wir in unserer Arbeit eine Kombination von NMR, EPR, SAXS und CD-Spektroskopie um die hoch flexible Beschaffenheit von OPN zu untersuchen. Unsere Resultate zeigen dass OPN sich durch ein vielschichtiges Strukturensemble in Lösung auszeichnet das ausgedehnte sowie kooperativ gefaltete Konformationen enthält. Schlussendlich untersuchen wir die Interaktion von OPN mit seinem natürlichen Ligand, Heparin.

Table of Contents

Introduction	11
Chapter I	25
Chapter II	61
Chapter III	97
Discussion	119
Acknowledgements	133
Curriculum Vitae	137

Introduction

Biological systems ranging from the most basic unicellular life forms to higher eukaryotic systems rely on a multitude of different building blocks (e.g. DNA, carbohydrates, proteins, lipids, *etc.*) that form the entity that is a living organism. Among all the components of these incredibly complex systems, proteins carry out a wide range of different functions. Many proteins exert their biological function by binding a ligand, which can either modulate its activity or be a substrate for catalysis. Published in 1894, Fischer's lock-and-key model¹ was the first mechanistic concept describing the molecular recognition between a protein and its ligand. In this model, the protein is treated as a static object. While this model was able to describe recognition and specificity it was not sufficient to explain the stabilization of transition states during enzymatic reactions. Sixty-four years later, in 1958, the "induced fit" model was postulated, which stated that a binding event can induce a conformational change in a protein.² This model, for the first time, included the concept of conformational dynamics in protein-ligand interaction. In 1999, the alternative concept of "conformational selection" arose which argued that protein conformations other than the "ground state" (lowest energy), commonly termed "excited states", often represent the relevant targets for recognition events.^{3,4} The model derives from the energy landscape theory of proteins published by Frauenfelder et al.⁵ In this new paradigm, suitable targets are recognized by low-populated, high-energy states of the protein, resulting in a subsequent population shift towards these states. To this day, Frauenfelder's "conformational selection" and Koshland's "induced fit" models are the two prevalent theories for protein-ligand interaction events. Today it is recognized that dynamics is crucial for protein function (e.g. allostery and molecular recognition). A prime example is seen in enzymes, which rely on motions taking place on a broad range of timescales ranging from bond vibrations (10-100 fs), to side chain motions (ps – ns) to conformational rearrangements of whole domains (μ s – ms) for catalysis.

The dynamic behavior of intrinsically disordered proteins (IDPs) presents an extreme case in the realm of proteins because IDPs display no stable tertiary structure under physiological conditions but exists as a heterogeneous and dynamic ensemble of different conformational states.⁶⁻⁸ This property is a direct consequence of the unique amino acid composition of the protein. IDPs are

usually depleted in bulky hydrophobic side chains that are necessary for structure stabilizing interactions. Instead these proteins are enriched in basic and acidic amino acids, which are not conducive to stable tertiary structures.⁸ In contrast to structured proteins the folding landscape of IDPs does not possess a global energy minimum. Energetic barriers between local energy minima are small and can easily be overcome by thermal fluctuations.

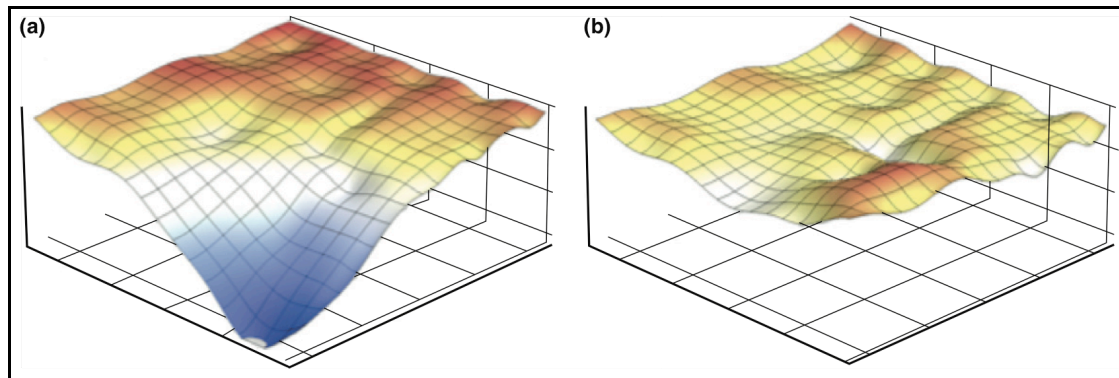


Figure 1: Energy landscapes of proteins. a) A folded protein exhibits a well-defined global energy minimum that corresponds to its stable folded conformation. b) The folding landscape of an IDP is lacking a global energy minimum and energy barriers between local minima are small. (Adapted from Fisher *et. al.*⁹)

The discovery of IDPs challenged the long prevalent notion that protein function is dependent on a well-defined tertiary structure and attracted a lot of attention in the scientific community. A lot of progress has been made in the bioinformatic prediction and identification of IDPs^{10,11} revealing their occurrence in all kingdoms of life with almost one-third of the eukaryotic proteome predicted to contain extended segments of disorder.^{12,13} Occurrence of disorder is even higher for proteins implicated in signaling as well as cancer-related diseases.^{14,15}

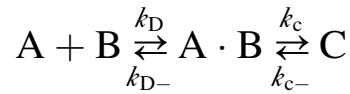
Structural disorder has very important implications from an evolutionary point of view. Gerhart *et. al.*¹⁶ proposed the theory of facilitated variation in which new phenotypes arise from different combinations of so-called conserved “core-components” (basic biochemical, genetic, cell biological, physiological, and developmental components) giving rise to so-called “core-functions” (energy metabolism, transcription, translation, etc.). Core-functions need to display several properties to be evolutionary relevant including modularity, robustness, adaptability, capacity to engage in weak regulatory linkage and exploratory

behavior.¹⁶ In a recent concept article by Konrat¹⁷ it was proposed that proteins exhibit all the important properties of such core-functions. With respect to the above-mentioned evolutionary relevant properties of such core-functions, proteins can achieve modularity by different combinations of their respective core-components (different secondary structure elements giving rise to tertiary structure, which in turn constitute multi-domain proteins). In the context of facilitated evolution regions of disorder or completely disordered proteins are of special interest. By exploiting conformational heterogeneity, modularity can be tuned to a large extent since the lack of structural predetermination allows for a large number of conformational states that can be selected by external factors. This structural freedom also endows flexible motifs with a high degree of adaptability, e.g., in response to external stimuli and different cellular environments. Additionally, regions of disorder are very robust and variations in disordered regions (e.g., mutations and alternative splicing) have only a marginal effect on stability and function with respect to stably folded proteins.¹⁸ Structural flexibility also has very important implications on signaling events. In protein interaction-networks, IDPs often play important regulatory roles and are able to act as “hub” proteins that signal to various targets ¹⁹ (e.g., p53, CBP, α -synuclein or BRCA1). By weakly linking other proteins (core-functions) IDPs allow for allosteric regulation and provide a way to tune the strength of protein interactions. As stated by Gerhart *et. al.*¹⁶, it is often a change in regulation of existing core-functions that triggers the emergence of new traits. Hence, the involvement of disordered regions in weak regulatory linkages between many different targets makes them a prime candidate for evolutionary alterations. A variation step (mutations or rearrangements of the genome) can result in links being formed or broken, or lead to an alteration of already existing linkages between different core-functions. This can ultimately result in the generation of new phenotypes. The high degree of conformational freedom thus provides disordered regions and IDPs respectively with a pronounced exploratory behavior priming them as a driving force of evolution.

The implications of disorder on the kinetics and thermodynamics of binding reactions are complex. Although models and theories on IDP binding have been postulated,²⁰⁻²³ a unifying theory has not yet emerged. While some

IDPs are able to engage with their binding partners in dynamic complexes, where the protein remains disordered in the bound state^{24,25}, most IDPs have been shown to fold into stable structures upon binding, experiencing a “disorder to order” transition in the process. This mechanism has been termed “folding upon binding” or “coupled binding and folding”²⁶⁻²⁸ and can proceed via two mechanisms. The interaction can take place with a suitable conformer that is in equilibrium with the unstructured ensemble (conformational selection)²⁹ or by binding of an unfolded state and subsequent folding (induced fit).³⁰ A clear distinction between these two extreme models however is not absolute and coupled folding and binding is thought to follow parallel pathways depending on the concentration of both IDP and its interaction partner.^{31,32} Several excellent papers address the existing ambiguities on both the binding modes of IDPs^{20,21,33}, as well as their thermodynamic and kinetic properties.^{34,35}

In general, binding of two proteins can be described by a two-step process represented by the following reaction scheme.



Here, A and B represent the two interacting proteins in the free state. The first step involves the formation of the encounter complex [A·B] whose rate of formation limited by rotational and translational diffusion (i.e. is diffusion limited) to bring the two interaction partners in close proximity. The second step comprises conformational rearrangements that lead to the final bound state [C]. Both the diffusion-limited step that leads to formation of the encounter complex and the rearrangement step can be rate-limiting. If the structural difference between the free and the bound state is large, the conformational step is regarded rate-limiting. It should be noted that in the conformation limited regime $k_c \ll k_D$ holds, while in the diffusion limited regime $k_c \gg k_D$. This results in association rates for protein-protein binding events that span a range of over 10 orders of magnitude ranging from $10^{10} \text{ M}^{-1}\text{s}^{-1}$ to $1 \text{ M}^{-1}\text{s}^{-1}$. While the theoretical upper limit for the association rate is set by diffusion to $\sim 10^5 \text{ M}^{-1}\text{s}^{-1}$, electrostatic

complementary binding surfaces can increase on-rates by orders of magnitude reaching k_D values of up to $10^{10} \text{ M}^{-1}\text{s}^{-1}$.³⁶

It has been argued that disorder can increase the on-rates in protein association events.³⁷ Shoemaker *et al.* proposed that an increased capture radius allows for a more efficient sampling of space for interaction partners by a mechanism termed fly-casting.³⁸ This mechanism has been critically revisited in a recent paper by Huang and Liu which compared binding kinetics of IDPs with ordered proteins.³⁹ In their study k_{on} and k_{off} rates of both IDPs and ordered proteins cover a range of 8 orders of magnitude. The authors were able to show that IDPs generally exhibit only moderately faster (~ 2.9 times) on and off rates. Furthermore, they state when taking the decreased diffusion rates (larger hydrodynamic radii) of IDPs into account, the benefit of a larger capture radius is almost eliminated. Subsequently, another more likely mechanism was proposed by Qin *et al.* In this dock-and-coalesce model, the IDP docks to its cognate binding partner in an initial encounter complex.⁴⁰ The structural plasticity allows the protein to sample the conformational space in this initial complex and coalesce onto the target. Here disorder is essential to overcome the orientational restraints that a folded protein would experience. Given the fact that many IDPs are maintained at low cellular concentrations⁴¹, this mechanism increases the likelihood to form the bound state with fewer encounter events needed, in contrast to folded proteins.

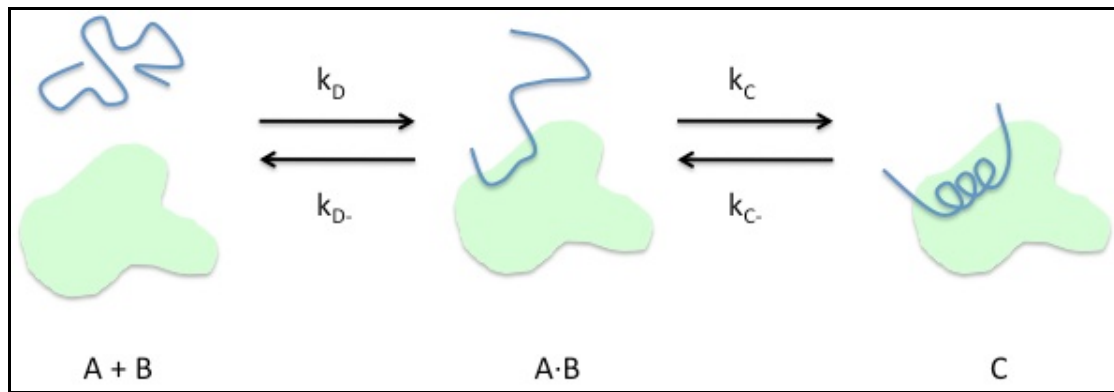


Figure 2: Schematic representation of the dock and coalesce model. In the first step (rate constant k_D) a segment of the IDP engages in an initial encounter complex $[A \cdot B]$ with its target. The IDP can subsequently either dissociate (rate constant k_{D-}) or coalesce onto the target (rate constant k_C) to form the final bound state $[C]$.

Considering the small difference (~ 2.9 times)³⁹ in kinetic rates of complex formation when comparing folded and disordered proteins, it can be reasoned that disorder itself does not impose a significant bias on the kinetics of binding since the same ranges of k_{on} , k_{off} and K_d values can be achieved by both types of proteins. Instead, disorder is likely to provide the means to tune interaction processes to fit cellular needs. This can, e.g., be achieved by utilizing PTMs to alter the net charge of an IDP, for example. This allows for a convenient way to alter kinetic rates and the lifetime of an interaction process. Additionally, PTMs are involved in targeting proteins for degradation as seen in phosphorylation-mediated ubiquitination of transcription factors.⁴² Indeed, IDP concentrations are usually maintained at much lower levels than structured proteins and a high concentration can lead to pathological effects.⁴³ The abundance of IDPs is tightly controlled by targeted protein degradation, which is also facilitated by their open structure.⁴⁴ PTMs can also shift the conformational equilibrium towards a specific conformational sub-ensemble. This can result in more enhanced desired or undesired interactions through stabilization of specific conformations. A relevant example is the formation of amyloid fibres in neurodegenerative diseases like Parkinson and Alzheimer.⁴⁵

Osteopontin (OPN) is a member of the intrinsically disordered protein class. It is a highly negatively charged protein, and a target of several post translational modifications, including phosphorylation and glycosylation.⁴⁶ The primary sequence of OPN is devoid of cysteines and depleted in bulky

hydrophobic amino acids, a common characteristic among IDPs.⁸ OPN was first indentified in rat bone by demineralization and subsequent purification of the tissue.⁴⁷ It is a member of the small integrin binding ligand, N-linked glycoprotein (SIBLING) family of proteins, along with Dentin Sialophosphoprotein (DSPP), Dentin Matrix Protein 1 (DMP1), Bone Sialoprotein (BSP) and Matrix Extracellular Phosphoglycoprotein (MEPE).⁴⁸ The group of Fedarko named this family of proteins after several genetic and biochemical features they have in common and were also the first to utilize NMR (nuclear magnetic resonance) to demonstrate their conformational flexibility.⁴⁹ OPN is involved in a multitude of physiological and pathological processes and a variety of different tissues and cell types express OPN.⁵⁰ OPN constitutes a major fraction among other non-collagenous proteins present in the extracellular matrix (ECM) of bone, cementum and dentin.⁵¹ After secretion into the extracellular space the protein interacts with hydroxyapatite thereby controlling calcium mineralization and the attachment of cells implicated in remodeling processes (e.g. osteoblasts), to the matrix. Cross-linking of OPN to other ECM-proteins, such as collagen, *via* transglutaminase enzymes can also add mechanical stability to the extracellular matrix.⁵² Furthermore, OPN is expressed in activated immune cells such as macrophages, T-cells and B-cells, acting as a cytokine and guiding these cells to sites of inflammation or injury. Epithelial cells (found in breast, urinary tract, gall bladder tissues) secrete OPN into biological fluids like blood, urine and milk. After secretion, OPN can also localize to the luminal surfaces of epithelial cells suggesting a protective role against unwanted interactions with the environment.⁵³ Finally, OPN is also found to be an inhibitor of calcium oxalate formation, the main crystalline component in kidney stones.⁵⁴

Besides its many physiological roles, OPN is overexpressed in a variety of malignant tissues.⁵⁵ Increased serum levels of OPN correlate with tumor grade and a poor survival rate of affected patients.⁵⁶ The protein exerts its tumorigenic function by interacting with CD44 receptors and several integrin variants through RGD-dependent (Arginine-Glycine-Aspartic Acid) and independent mechanisms. This interaction can alter cell-signaling events responsible for survival, motility and adhesion that ultimately lead to the formation of metastases. Consequently, OPN is found to be highly involved in all stages of

tumor progression: invasion, metastasis and angiogenesis.^{57,58} Pre- (alternative splicing) and post-translational modifications such as phosphorylation, glycosylation and proteolytic cleavage have been found to be cell-type specific and might influence the malignant potential of the protein.⁵⁹ The interaction with matrix metalloproteinases (MMPs) in the process of ECM degradation and remodeling is of special importance during cancer invasion.⁶⁰

So far, many biological functions of IDPs have been revealed by the wealth of available molecular biological tools. Mutational analysis in living organisms coupled with cell biological and immunological experiments provide a general picture of the interaction networks that IDPs form. Their mode of action however, is poorly understood, especially at the atomic level. Furthermore, studying disordered states and rapidly inter-converting species poses a methodological challenge for conventional structure determination by X-ray crystallography. Instead, several solution-based methods find application in the study of IDPs. Circular dichroism (CD) can be used to estimate the secondary structure content present in the structural ensemble.⁶¹ Small angle X-ray scattering (SAXS), and other hydrodynamic experiments, estimate the overall shape and distribution of distances present in the protein.⁶² Fluorescence-based methods like Förster resonance energy transfer (FRET) yields distance distributions between two fluorophore moieties introduced via site-directed labeling.⁶³ Finally, continuous wave - electron paramagnetic enhancement (CW EPR) reports on local spin-label mobility of MTSL-labeled sites. NMR is arguably the most comprehensive biophysical tool since it is able to give a highly detailed picture of the dynamic properties of IDPs that accompany their biological activity. Information about local dynamics (spin relaxation), transient contact formation (PRE-paramagnetic relaxation enhancement), solvent exposure (hydrogen exchange), backbone dihedral angles (scalar couplings), residual secondary structure propensities (secondary chemical shift mapping) and orientation of bond vectors with respect to the external field (RDC-residual dipolar couplings) can be obtained with atomic resolution. Most importantly, NMR experiments cover a wide spectrum of timescales accessible to proteins probing motions in the range of picoseconds to seconds and higher.⁶⁴ In particular, CPMG (Carr-Purcell-Meiboom-Gill) relaxation dispersion experiments

are a promising tool for detecting weakly populated states, provided these states are adequately populated ($>1\%$) and the inter-conversion rate between them is in a suitable motional-window (ms- μ s).⁶⁵ Presently, the development of analytical procedures that are able to transform experimental observables of IDPs into meaningful ensemble descriptions is an important challenge.^{9,66,67} Two approaches are most widely used: The first one are molecular dynamics (MD) based approaches, combining experimental restraints with energy force field calculations to generate ensembles satisfying the experimental data. The second approach uses algorithms to initially seed the conformational space and subsequently populate and select structures that satisfy the experimental data. Regardless of the approach used, the result can only provide ensemble-averaged data. When dealing with a highly heterogeneous conformational space, as is the case for IDPs, information about individual sub-conformers is lost in the averaging process if the rate of conversion between states takes place on timescales shorter than the temporal resolution of the applied technique.

In this work we use a large variety of biophysical techniques in an integrated structural biology approach to characterize the conformational plasticity of the IDP Osteopontin. In **Chapter I** we show, using a combination of multidimensional NMR, SAXS and synchrotron radiation circular dichroism (SRCD) that OPN is lacking any stable tertiary structure elements. Instead it exhibits transient secondary structure formation along the protein backbone in regions where ligand interaction takes place. Using PRE we investigate the structural ensemble revealing transient long-range interactions that are important for the stabilization of OPNs compact sub-structures. In **Chapter II** we use a combination of EPR and NMR to combine dynamic data (from NMR) and static data double electron-electron resonance (DEER) - EPR to draw a detailed picture of the transition events taking place in the structural ensemble of OPN. Furthermore we observe the existence of cooperatively folded sub-structures present in the conformational ensemble. In **Chapter III** we thoroughly examine the ligand interaction properties of OPN with its target heparin and establish a detailed description of the molecular driving forces behind the binding process.

REFERENCES

1. Fischer, E., Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894**, 27 (3), 2985–2993.
2. Koshland, D. E., Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **1958**, 44 (2), 98-104.
3. Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R., Folding funnels and binding mechanisms. *Protein Eng* **1999**, 12 (9), 713-20.
4. Tsai, C. J.; Kumar, S.; Ma, B.; Nussinov, R., Folding funnels, binding funnels, and protein function. *Protein Sci* **1999**, 8 (6), 1181-90.
5. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G., The energy landscapes and motions of proteins. *Science* **1991**, 254 (5038), 1598-603.
6. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **1999**, 293 (2), 321-31.
7. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., Intrinsically disordered protein. *J Mol Graph Model* **2001**, 19 (1), 26-59.
8. Tompa, P., Intrinsically unstructured proteins. *Trends Biochem Sci* **2002**, 27 (10), 527-33.
9. Fisher, C. K.; Stultz, C. M., Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol* **2011**, 21 (3), 426-31.
10. Bracken, C.; Iakoucheva, L. M.; Romero, P. R.; Dunker, A. K., Combining prediction, computation and experiment for the characterization of protein disorder. *Curr Opin Struct Biol* **2004**, 14 (5), 570-6.
11. Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B., Protein disorder prediction: implications for structural proteomics. *Structure* **2003**, 11 (11), 1453-9.
12. 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* **2004**, 337 (3), 635-45.
13. Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J., Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* **2000**, 11, 161-71.
14. Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K., Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **2002**, 323 (3), 573-84.
15. Pajkos, M.; Meszaros, B.; Simon, I.; Dosztanyi, Z., Is there a biological cost of protein disorder? Analysis of cancer-associated mutations. *Mol Biosyst* **2012**, 8 (1), 296-307.
16. Gerhart, J.; Kirschner, M., The theory of facilitated variation. *Proc Natl Acad Sci U S A* **2007**, 104 Suppl 1, 8582-9.
17. Konrat, R., The meandering of disordered proteins in conformational space. *Structure* **2010**, 18 (4), 416-9.
18. Brown, C. J.; Takayama, S.; Campen, A. M.; Vise, P.; Marshall, T. W.; Oldfield, C. J.; Williams, C. J.; Dunker, A. K., Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol* **2002**, 55 (1), 104-10.

19. Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N., Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **2005**, 272 (20), 5129-48.
20. Csermely, P.; Palotai, R.; Nussinov, R., Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **2010**, 35 (10), 539-46.
21. Kiefhaber, T.; Bachmann, A.; Jensen, K. S., Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr Opin Struct Biol* **2012**, 22 (1), 21-9.
22. 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* **2012**, 8 (1), e1002353.
23. Ganguly, D.; Zhang, W.; Chen, J., Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. *Mol Biosyst* **2012**, 8 (1), 198-209.
24. Sigalov, A. B.; Zhuravleva, A. V.; Orekhov, V. Y., Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form. *Biochimie* **2007**, 89 (3), 419-21.
25. Mittag, T.; Orlicky, S.; Choy, W. Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D., Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci U S A* **2008**, 105 (46), 17772-7.
26. Dyson, H. J.; Wright, P. E., Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **2002**, 12 (1), 54-60.
27. Sugase, K.; Dyson, H. J.; Wright, P. E., Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, 447 (7147), 1021-5.
28. Wright, P. E.; Dyson, H. J., Linking folding and binding. *Curr Opin Struct Biol* **2009**, 19 (1), 31-8.
29. Monod, J.; Wyman, J.; Changeux, J. P., On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol* **1965**, 12, 88-118.
30. Koshland, D. E., Jr.; Nemethy, G.; Filmer, D., Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **1966**, 5 (1), 365-85.
31. Weikl, T. R.; von Deuster, C., Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. *Proteins* **2009**, 75 (1), 104-10.
32. Hammes, G. G.; Chang, Y. C.; Oas, T. G., Conformational selection or induced fit: a flux description of reaction mechanism. *Proc Natl Acad Sci U S A* **2009**, 106 (33), 13737-41.
33. Chen, J., Towards the physical basis of how intrinsic disorder mediates protein function. *Arch Biochem Biophys* **2012**, 524 (2), 123-31.
34. Zhou, H. X., Intrinsic disorder: signaling via highly specific but short-lived association. *Trends Biochem Sci* **2012**, 37 (2), 43-8.
35. Mittag, T.; Kay, L. E.; Forman-Kay, J. D., Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* **2009**, 23 (2), 105-16.
36. Schreiber, G.; Haran, G.; Zhou, H. X., Fundamental aspects of protein-protein association kinetics. *Chem Rev* **2009**, 109 (3), 839-60.

37. Pontius, B. W., Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem Sci* **1993**, *18* (5), 181-6.
38. 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* **2000**, *97* (16), 8868-73.
39. Huang, Y.; Liu, Z., Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J Mol Biol* **2009**, *393* (5), 1143-59.
40. Qin, S.; Pang, X.; Zhou, H. X., Automated prediction of protein association rate constants. *Structure* **2011**, *19* (12), 1744-51.
41. Gsponer, J.; Futschik, M. E.; Teichmann, S. A.; Babu, M. M., Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* **2008**, *322* (5906), 1365-8.
42. Karin, M.; Ben-Neriah, Y., Phosphorylation meets ubiquitination: the control of NF- κ B activity. *Annu Rev Immunol* **2000**, *18*, 621-63.
43. Vavouri, T.; Semple, J. I.; Garcia-Verdugo, R.; Lehner, B., Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **2009**, *138* (1), 198-208.
44. Tsvetkov, P.; Reuven, N.; Shaul, Y., The nanny model for IDPs. *Nat Chem Biol* **2009**, *5* (11), 778-81.
45. Chiti, F.; Dobson, C. M., Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **2006**, *75*, 333-66.
46. 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* **2005**, *390* (Pt 1), 285-92.
47. Prince, C. W.; Oosawa, T.; Butler, W. T.; Tomana, M.; Bhown, A. S.; Bhown, M.; Schrohenloher, R. E., Isolation, characterization, and biosynthesis of a phosphorylated glycoprotein from rat bone. *J Biol Chem* **1987**, *262* (6), 2900-7.
48. Bellahcene, A.; Castronovo, V.; Ogbureke, K. U.; Fisher, L. W.; Fedarko, N. S., Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer. *Nat Rev Cancer* **2008**, *8* (3), 212-26.
49. Fisher, L. W.; Torchia, D. A.; Fohr, B.; Young, M. F.; Fedarko, N. S., Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochem Biophys Res Commun* **2001**, *280* (2), 460-5.
50. Kunii, Y.; Niwa, S.; Hagiwara, Y.; Maeda, M.; Seitoh, T.; Suzuki, T., 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* **2009**, *42* (3), 155-61.
51. McKee, M. D.; Nanci, A., Osteopontin: an interfacial extracellular matrix protein in mineralized tissues. *Connect Tissue Res* **1996**, *35* (1-4), 197-205.
52. Kaartinen, M. T.; Pirhonen, A.; Linnala-Kankkunen, A.; Maenpaa, P. H., Cross-linking of osteopontin by tissue transglutaminase increases its collagen binding properties. *J Biol Chem* **1999**, *274* (3), 1729-35.
53. Brown, L. F.; Berse, B.; Van de Water, L.; Papadopoulos-Sergiou, A.; Perruzzi, C. A.; Manseau, E. J.; Dvorak, H. F.; Senger, D. R., Expression and

distribution of osteopontin in human tissues: widespread association with luminal epithelial surfaces. *Mol Biol Cell* **1992**, 3 (10), 1169-80.

54. Wesson, J. A.; Johnson, R. J.; Mazzali, M.; Beshensky, A. M.; Stietz, S.; Giachelli, C.; Liaw, L.; Alpers, C. E.; Couser, W. G.; Kleinman, J. G.; Hughes, J., Osteopontin is a critical inhibitor of calcium oxalate crystal formation and retention in renal tubules. *J Am Soc Nephrol* **2003**, 14 (1), 139-47.

55. Brown, L. F.; Papadopoulos-Sergiou, A.; Berse, B.; Manseau, E. J.; Tognazzi, K.; Perruzzi, C. A.; Dvorak, H. F.; Senger, D. R., Osteopontin expression and distribution in human carcinomas. *Am J Pathol* **1994**, 145 (3), 610-23.

56. Weber, G. F.; Lett, G. S.; Haubein, N. C., Osteopontin is a marker for cancer aggressiveness and patient survival. *Br J Cancer* **2010**, 103 (6), 861-9.

57. 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* **2007**, 16 (6), 1087-97.

58. Wai, P. Y.; Kuo, P. C., The role of Osteopontin in tumor metastasis. *J Surg Res* **2004**, 121 (2), 228-41.

59. Anborgh, P. H.; Mutrie, J. C.; Tuck, A. B.; Chambers, A. F., Pre- and post-translational regulation of osteopontin in cancer. *J Cell Commun Signal* **2011**, 5 (2), 111-22.

60. Agnihotri, R.; Crawford, H. C.; Haro, H.; Matrisian, L. M.; Havrda, M. C.; Liaw, L., Osteopontin, a novel substrate for matrix metalloproteinase-3 (stromelysin-1) and matrix metalloproteinase-7 (matrilysin). *J Biol Chem* **2001**, 276 (30), 28261-7.

61. Chemes, L. B.; Alonso, L. G.; Noval, M. G.; de Prat-Gay, G., Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol Biol* **2012**, 895, 387-404.

62. Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I., Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* **2007**, 129 (17), 5656-64.

63. Ferreon, A. C.; Moran, C. R.; Gambin, Y.; Deniz, A. A., Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol* **2010**, 472, 179-204.

64. Mittermaier, A. K.; Kay, L. E., Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* **2009**, 34 (12), 601-11.

65. Tollinger, M.; Kloiber, K.; Agoston, B.; Dorigoni, C.; Lichtenegger, R.; Schmid, W.; Konrat, R., An isolated helix persists in a sparsely populated form of KIX under native conditions. *Biochemistry* **2006**, 45 (29), 8885-93.

66. Jensen, M. R.; Ruigrok, R. W.; Blackledge, M., Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* **2013**, 23 (3), 426-35.

67. Mittag, T.; Forman-Kay, J. D., Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **2007**, 17 (1), 3-14.

Chapter I

The Metastasis-Associated Extracellular Matrix Protein Osteopontin Forms Transient Structure in Ligand Interaction Sites

Platzer G, Schedlbauer A, Chemelli A, Ozdowy P, Coudevylle N, Auer R, Kontaxis G, Hartl M, Miles AJ, Wallace BA, Glatter O, Bister K, Konrat R.

Biochemistry **2012** 50(27), 6113-6124

ABSTRACT

Osteopontin (OPN) is an acidic hydrophilic glycoposphoprotein that was first identified as a major sialoprotein in bones. It functions as a cell attachment protein displaying a RGD cell adhesion sequence and as a cytokine that signals through integrin and CD44 cell adhesion molecules. OPN is also implicated in human tumor progression and cell invasion. OPN has intrinsic transforming activity, and elevated OPN levels promote metastasis. OPN gene expression is also strongly activated in avian fibroblasts simultaneously transformed by the *v-myc* and *v-mil(raf)* oncogenes. Here we have investigated the solution structure of a 220-amino acid recombinant OPN protein by an integrated structural biology approach employing bioinformatic sequence analysis, multi-dimensional nuclear magnetic resonance spectroscopy, synchrotron radiation circular dichroism spectroscopy, and small-angle X-ray scattering. These studies suggest that OPN is an intrinsically unstructured protein in solution. Although OPN does not fold into a single defined structure, its conformational flexibility significantly structure elements with reduced conformational flexibility and substantially populates a compact subspace displaying distinct tertiary contacts. These compacted regions of OPN encompass the binding sites for $\alpha_v\beta_{III}$ integrin and heparin. The conformational flexibility combined with the modular architecture of OPN may represent an important structural prerequisite for its functional diversity.

INTRODUCTION

Osteopontin (OPN), a secreted, sialic acid-rich, chemokine-like protein, is a member of the SIBLING family (small integrin-binding ligand N-linked glycoproteins).¹ OPN and other members of this protein family like bone sialoprotein, dentin matrix protein 1, dentin sialophosphoprotein, and matrix extra-cellular phosphoglycoprotein play important roles in many stages of cancer progression, as identified by gene transfer experiments.^{2,3} OPN and most of the other family members are strongly over-expressed in several cancer cell lines, and their abundance directly correlates with the tumor grade.⁴ For example, elevated expression levels of OPN were found in avian fibroblasts

cooperatively transformed by the v-myc and v-mil(raf) oncogenes encoded by the acute leukemia and carcinoma virus MH2.⁵ Promoter analysis revealed that the OPN gene is a direct transcriptional target gene of the activator protein-1 (AP-1) complex, mediated by a nonconsensus AP-1 binding site.⁵ Overexpression of OPN results in an increased level of malignancy, whereas RNA silencing using appropriate antisense oligonucleotides leads to less malignant phenotypes.⁶ A hallmark of OPN's biological functionality is the controlled interaction with a diverse set of receptor proteins.⁷ OPN exerts its regulatory effects predominantly by interaction with integrin and CD44 receptors.⁷ Additionally, OPN and other members of the SIBLING family have the capacity to bind and to activate matrix metalloproteases (MMPs).¹ The roles of OPN in diverse biological processes like atherosclerosis, bone remodeling, angiogenesis, wound healing, and tissue injuries have been reported.⁷⁻⁹ The involvement of OPN in certain diseases such as arterial hyperplasia, myocardial necrosis, renal fibrosis, and autoimmune diseases is well-documented.¹⁰ Of utmost medical importance is the fact that OPN is implicated in the pathology of diverse physiological processes as well as in tumor progression and metastasis in various types of cancer.^{1,7} OPN is classified as an arginine-glycine-aspartate (RGD)-containing glycoprotein predominantly found in kidney, brain, macrophages, vascular smooth muscle, and epithelial cells. It is an acidic protein, rich in aspartic and glutamic acids, and highly glycosylated, having an average molecular mass of ~44 kDa. The biological functionality of OPN is regulated by extensive phosphorylation.¹¹ It has a protease cleavage site separating two protein protein interaction subdomains, an integrin-binding motif comprising the RGD sequence, and a C-terminal CD44 receptor binding site. However, the *in vivo* relevance of OPN activation by thrombin cleavage is still unclear. Finally, OPN also serves as a substrate for both liver transglutaminase and plasma transglutaminase factor XIIIa. Comprehensive overviews of OPN's biological functions of OPN have been published in recent years.^{2,3,7}

Despite the significant biological and biomedical relevance of OPN, detailed information about its solution structural properties is still largely missing. Some preliminary insight was provided by recent NMR studies, where local structuring of OPN was demonstrated.^{12,13} Here we have investigated the

solution structure of a recombinant OPN protein by an integrated structural biology approach employing multidimensional nuclear magnetic resonance (NMR) spectroscopy, synchrotron radiation circular dichroism (SRCD) spectroscopy, and small-angle X-ray scattering (SAXS). We show that OPN is devoid of any significant tertiary structure, largely lacking secondary structures, and thus represents another member of the growing family of intrinsically disordered and conformationally flexible proteins. The conformational ensemble of quail OPN was further investigated using residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs). For the identification of suitable sites of spin-label attachment, we employed our recently developed meta-structure approach.¹⁴ NMR spectroscopy was also used to investigate putative calcium ion and heparin binding sites in the OPN protein. No evidence of calcium binding was obtained, but the binding site of heparin was mapped to the central integrin-binding domain of OPN comprising the RGD motif. On the basis of multidimensional NMR spectroscopy, we provide evidence that OPN exists as a conformationally flexible polypeptide chain in solution devoid of stabilizing long-range interactions. However, despite the fact that OPN does not form an energetically well-defined three-dimensional (3D) solution structure, it displays several local secondary structure elements with significantly reduced conformational flexibilities. Additionally, SAXS and PRE data indicate that although OPN exists in a rather extended form, there is evidence of a transient compaction of the polypeptide chain. This motional heterogeneity along the protein backbone may be of relevance for the biological function of OPN. It could allow adaptive processes for optimization and tuning of intermolecular interactions with diverse protein binding partners across the relevant interaction interfaces of OPN.

MATERIALS AND METHODS

Protein Expression

The recombinant quail OPN protein (OPN220) contains the 219 C-terminal amino acids of quail OPN and a vector-encoded start methionine.^{5,12} *Escherichia coli* strain BL21(DE3) transformed with the pET11d-OPN220

expression vector described previously¹² was grown in LB medium overnight. The culture was diluted 1:100 in 50 mL of minimal medium, supplemented with 0.5 g of NaCl, 1 g of $^{15}\text{NH}_4\text{Cl}$, and 2 g of D- ^{13}C glucose per 1000 mL, and the bacteria were incubated at 37 °C for 12-16 h. The 50 mL preculture was added to 1000 mL of minimal medium, and the bacteria were grown at 37 °C until the culture reached an absorbance at 600 nm of 0.7. Recombinant protein synthesis was induced by the addition of isopropyl β -D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and incubation was continued for 12 h at 28 °C. The cells were collected by centrifugation at 5000 rpm for 20 min and resuspended in 20 mL of ice-cold PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , and 1.8 mM KH_2PO_4) per liter of the original bacterial culture. Bacteria were lysed by sonication, and the cell lysate was cleared by centrifugation at 18000 rpm for 30 min; 20 mL portions of the supernatant containing the soluble protein fraction were adjusted to a $(\text{NH}_4)_2\text{SO}_4$ concentration of 50% and then clarified by centrifugation at 18000 rpm for 30 min. The precipitate containing most of the OPN recombinant protein was dialyzed against PBS buffer at 4 °C and loaded onto a Resource Q, 6 mL ion exchange column (GE Healthcare). Protein was eluted with a sodium chloride gradient generated by mixing PBS and high-salt PBS buffer (1 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , and 1.8 mM KH_2PO_4). The fractions containing most of the OPN protein were concentrated to 1.5 mL by centrifugation through an Amicon Ultra-15 centrifugal filter device 10K NMWL (Amicon) and loaded onto a Superdex 200 HiLoad 16/60 prep grade (GE Healthcare) gel filtration column equilibrated in phosphate buffer [50 mM NaCl and 50 mM sodium phosphate (pH 6.5)]. The final yield of homogeneous OPN was approximately 4 mg/L of bacterial culture. For NMR analysis, protein samples were concentrated to 0.5 mM.

Synchrotron Radiation Circular Dichroism (SRCD) Spectroscopy

Protein concentrations were determined by absorbance measurements (NanoDrop UV/vis spectrophotometer using the theoretical extinction coefficient calculated using the ProtParam tool at <http://www.expasy.org>). Samples containing 0.5 mM of the human form of the protein dissolved in phosphate buffer [50 mM NaCl and 50 mM sodium phosphate (pH 6.5)] were

loaded into a demountable Suprasil quartz cell (Hellma UK Ltd.) with a path length of 0.0015 cm. The spectrum was measured at 25 °C over the wavelength range from 170 to 280 nm with a step size of 1 nm and a dwell time of 2 s. Data were processed using CDTool.¹⁵ Three sample scans were averaged, and three baseline scans (consisting of buffer) were averaged and subtracted from the sample scans. The resulting spectrum was smoothed using a Savitsky-Golay filter, calibrated against a spectrum of camphorsulfonic acid,¹⁶ and then converted to $\Delta\epsilon$ units using a mean residue weight of 113.7. Secondary structure content was analyzed using the DichroWeb analysis server,¹⁷ using reference data set 6,¹⁸ which contains the spectra of five denatured proteins and hence produces a more accurate prediction when the protein contains disordered structure. The results from three different algorithms, CONTINLL,^{19,20} SELCON3,²¹ and CDSSTR,²¹ were averaged.

NMR Spectroscopy

NMR spectra were recorded at 25 °C using Varian Inova 500 and 800 MHz and Varian Direct Drive 600 MHz spectrometers. Spectra were recorded in the States-TPPI/PFG sensitivity-enhanced mode²² for quadrature detection with carrier frequencies for $^1\text{H}^{\text{N}}$ and ^{15}N of 4.73 and 120.0 ppm, respectively. The sample contained 0.5 mM recombinant quail OPN, 50 mM NaCl, and 50 mM sodium phosphate (pH 6.5) in a 90% H_2O /10% D_2O mixture. The suite of standard triple-resonance experiments performed for extraction of $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and $^{13}\text{C}'$ chemical shifts included 3D HNCA, 3D HN(CO)CA, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO, and 3D HN(CA)CO.²³ In addition, 3D HNN and 3D HN(C)N experiments²⁴ were conducted to establish sequential connectivities and to obtain a large extent of unambiguous backbone assignments. NMR spectra were processed and analyzed with NMRPipe²⁵ and SPARKY.²⁶ High-quality NMR data for OPN were obtained as shown by the PFG sensitivity-enhanced two-dimensional (2D) ^1H - ^{15}N HSQC spectrum recorded at 800 MHz (cf. Figure 1c).

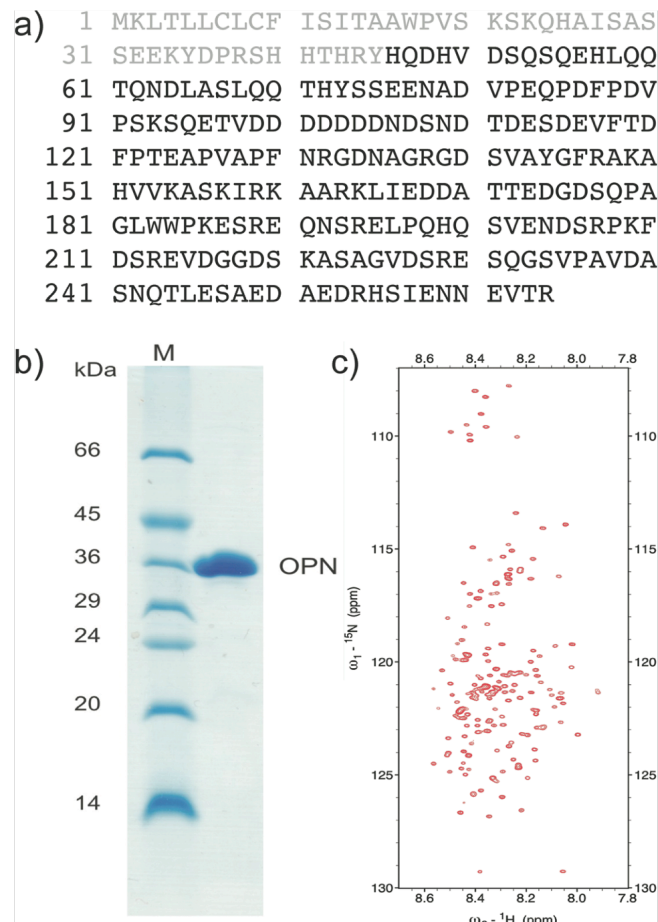


Figure 1. Purity of the 220-amino acid recombinant OPN (OPN220) monitored by sodium dodecyl sulfate-polyacrylamide (17%, w/v) gel electrophoresis and 2D ^1H ^{15}N heteronuclear single-quantum coherence (HSQC) spectroscopy. (a) Amino acid sequence of wild-type quail OPN. OPN220 contains the 219 carboxyl-terminal amino acids of OPN and an initiating methionine specified by the vector construct and lacks the 45 N-terminal residues (gray) of OPN. (b) OPN expressed and purified from the *E. coli* strain BL21(DE3) transformed with the pET11d-OPN220 expression vector after final gel filtration chromatography. Proteins were visualized after being stained with Coomassie brilliant blue. M, molecular weight markers. The OPN220 protein ($M_r=24311$; $pI=3.94$) displays an apparent molecular weight of 35000 (p35). The large deviation from the true M_r is due to the anomalous electrophoretic mobility in SDS-PAGE typically observed for many proteins with low pI values.⁵ (c) 2D ^1H - ^{15}N HSQC spectroscopy of recombinant OPN. The narrow dispersion of cross-peaks indicates the lack of a defined tertiary structure.

A squared and 60° phase-shifted sine bell window function was applied in all dimensions for apodization. Time domain data were zero-filled to twice the data set size, prior to Fourier transformation. The applied pulse sequences are

sensitivity-enhanced and use gradients for coherence selection and water suppression. ^1H chemical shifts were referenced as reported previously.¹² Heparin was purchased from Sigma Aldrich and used without further treatment.

Secondary Chemical Shifts

Using the assigned $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and $^{13}\text{C}'$ chemical shifts, the consensus chemical shift index (CSI) was employed to detect possible preferences for secondary structure elements.²⁷ Random coil values were subtracted from measured chemical shifts after correction for sequence effects as described previously.²⁸ For illustration, gliding averages along the amino acid positions were calculated using a window size of five residues.

^{15}N Relaxation

Dynamics information was obtained by measurement of backbone ^{15}N -T1 and -T2 relaxation times as described previously.²⁹ Relaxation parameters were measured at 25 °C employing a magnetic field strength that corresponds to a ^1H Larmor frequency of 599.89 MHz. The ^{15}N transverse relaxations experiments for evaluating T2 were performed with Carr Purcell Meiboom Gill (CPMG) delays of 0, 16, 32, 64, 128, 192, and 256 ms using a CPMG duty cycle delay of 0.5 ms. ^{15}N longitudinal relaxation T1 measurements were obtained by employing delays of 0, 20.8, 41.6, 83.2, 208, 416, 832, 1248, and 1664 ms. The sequence of cross-peak intensities in this series of spectra was fit (using an extension script of SPARKY²⁶) to a two-parameter exponential decay of the form $I(t) = A \exp[-(t/T)]$, where $I(t)$ represents the peak intensity and t the delay time. Random errors for T1 and T2 values were estimated by calculating a best fit for a set of perturbed heights 5000 times. The heights were perturbed by computer-simulated noise with a Gaussian distribution of zero mean and with variance equal to the root-mean-square deviation of the original heights from the original best fit. The spread in the best fit values was used as a measure of the experimental random error. Heteronuclear steady-state NOE $^{15}\text{N}\{-^1\text{HN}\}$ attenuation factors were derived from the $I_{\text{NOE}}/I_{\text{nonNOE}}$ ratio, where I_{NOE} and I_{nonNOE} denote the peak intensities in the experiments with and without proton saturation, respectively.²⁹ In the case of spectra with nonsaturation, a net

relaxation delay of 5 s was employed whereas a relaxation delay of 2 s prior to a 3 s proton presaturation period was applied for the NOE spectra. Acquisition parameters were identical to those of the PFG sensitivity-enhanced 2D ^1H - ^{15}N HSQC experiment except that the number of transients per t1 value was set to 32. ^{15}N CPMG dispersion experiments were performed as described by Kay and co-workers.³⁰

Measurement of Hydrodynamic Radii

Pulse field gradient (PFG) NMR diffusion measurements were performed on a Varian Direct Drive spectrometer equipped with a z-gradient probe operating at 600 MHz. The sample was supplemented with 0.2% dioxane as an internal reference. The PG-SLED (pulse gradient stimulated echo longitudinal encode-decode) sequence was used as described by Wilkins et al.,³¹ to which a final WATERGATE module³² was added. The applied diffusion gradient time (Δ) was 4.5 ms, and the echo time (δ) was 100 ms. Fifty experiments were conducted with increasing diffusion gradient strength. The data were processed using NMRPipe.²⁵ The decay of the protein aliphatic proton signals was fitted to a single Gaussian using the DOSY module of NMRPipe to determine the diffusion constant of the protein (d_{prot}). The same procedure was applied to the dioxane resonance to determine the diffusion constant of dioxane (d_{diox}). The hydrodynamic radius of the protein ($R_{\text{h}}^{\text{prot}}$) was deduced using the following equation:

$$R_{\text{h}}^{\text{prot}} = \frac{d_{\text{diox}}}{d_{\text{prot}}} (R_{\text{h}}^{\text{diox}})$$

assuming that $R_{\text{h}}^{\text{diox}}$ (hydrodynamic radius of dioxane) is 2.12 Å.³¹

Residual Dipolar Couplings (RDCs)

To obtain ^{15}N - ^1HN couplings, a ^1H - ^{15}N HSQC spectrum without proton decoupling during nitrogen evolution was recorded. The resulting signal splitting leads to a doubling of peaks, which might result in a crowded spectrum with signal overlap. To circumvent this problem, we recorded two separate spectra.

One contains the inphase component of the coupled magnetization and the other the antiphase component (IPAP). Both spectra were added and subtracted from each other to separately yield the upfield and downfield components of the doublet. Splittings were measured under isotropic (J_{HN}) and aligned ($J_{HN} + D_{HN}$) conditions to isolate the RDC contribution.³³

Paramagnetic Relaxation Enhancements (PREs)

Paramagnetic relaxation enhancements by nitroxide spin-labels [4-(2-iodoacetamido)-TEMPO] in single-cysteine-mutant OPN variants were measured as intensity ratios of cross-peaks in 2D ^1H - ^{15}N HSQC NMR spectra in the presence and absence of the nitroxide radical. The individual cross-peak intensities were normalized using the most intense peaks in the two data sets. The mutant OPN proteins were expressed and purified from plasmids that have been mutagenized using the original pET11d-OPN220 construct as a template and the QuickChange II site-directed mutagenesis kit (Stratagene). The following complementary (f, forward; r, reverse) primers were used:

Opn220-c108-f, 5'-GATGATGACAATGATTGCAATGACACCGATGAG-3';

Opn220-c108-r, 5'-CTCATCGGTGTCATTGCAATCATTGTTCATCATC-3';

Opn220-c188-f, 5'-TGGTGGCCCAAGGAGTGCCGCGAACAGAACAGC3';

Opn220-c188-r, 5'-GCTGTTC- TGTTCGCGGCACTCCTTGGGCCACCA-3'.

All mutant constructs were verified by DNA sequencing. Cysteine-containing OPN mutant proteins were expressed and purified as described for wild-type OPN except that dithiothreitol (DTT) was present at a final concentration of 1 mM in all buffers used in the purification procedure.

Small-Angle X-ray Scattering

The SAXS equipment consisted of a slit-geometry camera with high flux and low back- ground (SAXSess, Anton-Paar, Austria) connected to an X-ray generator (Philips, PW1730/10) operating at 40 kV and 50 mA with a sealed-tube Cu anode ($\lambda = 0.154$ nm). A CCD camera (Princeton Instruments/Roper Scientifics, Trenton, NJ) was used to record the 2D scattering pattern. The temperature of the capillary in the metallic sample holder was controlled by a Peltier element. Measurements were taken at 25 °C. The measuring times were 3

x 10 min for all samples. This allowed for the proper subtraction of cosmic rays that appear when using a CCD camera. The images were then integrated into the one-dimensional scattering function $I(q)$. This one-dimensional scattering curve is a function of the magnitude of the scattering vector $q = (4\pi/\lambda) \sin(\theta/2)$, where θ is the total scattering angle. All the intensities were transmission-calibrated by normalizing the attenuated primary intensity at $q = 0$ to unity and were corrected for the background scattering from the capillary and the buffer solution. The absolute scale calibration was achieved by using water as a secondary standard.³⁴

The obtained SAXS spectra were further analyzed with indirect Fourier transformation (IFT).³⁵ IFT is a model-free method used for dilute particle systems with negligible particle interactions. In dilute solutions, interparticle interactions can be neglected. The scattering intensity from the scattering particles $I(q)$ can be written as the Fourier transformation of the so-called pair-distance distribution function (PDDF) $p(r)$ describing the geometry of the scattering particles in real space:³⁶

$$I_1(q) = 4\pi \int_0^\infty p(r) \frac{\sin(qr)}{qr} dr$$

where r is the distance between two scattering centers within the particle. The $p(r)$ function represents a histogram of the distance inside the scattering particle,³⁶ which means that the function adopts a value of zero at a distance r greater than the maximum dimension of the particle. Particle size and shape are two important parameters that can be obtained from the $p(r)$ function.

RESULTS

NMR Spectroscopic Analysis of OPN and Resonance Assignments

Because of inefficient bacterial expression of the full-length quail OPN protein, a 220-amino acid recombinant OPN derivative lacking amino acid residues 2-45 of the 264-amino acid quail OPN was expressed and purified as described previously.¹² It should be noted that all amino acid positions

addressed in this report refer to the numbering of the full-length 264-amino acid protein. The purity of the protein sample was verified by SDS PAGE and NMR spectroscopy (Figure 1). The acidic OPN220 ($M_r = 24311$; $pI = 3.94$) migrates with an apparent molecular weight of 35000 (Figure 1b). This anomalous electrophoretic behavior in SDS PAGE is typical for highly acidic proteins.⁵ Electrospray ionization (ESI) mass spectrometry was performed to confirm the identity of the purified recombinant protein (data not shown). The ^1H - ^{15}N HSQC spectrum shows a distinct separation of the cross-peaks into regions typical of residue random coil positions (Figure 1c). There are resonance clusters for glycines, serines/threonines, and all other backbone amides. One underlying problem in the assignment of an intrinsically unfolded protein like OPN is the rather poor extent of backbone amide ^1HN , $^{13}C\alpha$, and $^{13}C\beta$ chemical shift dispersion compared to that of a protein molecule in the folded state. Ring current and other dipolar interactions in unfolded states are averaged out by conformational fluctuations leading to chemical shifts close to random coil values. Consequently, the ^1HN , $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts for the same residue type tend to degenerate. Furthermore, considering the relatively large molecular size of the 220-amino acid OPN protein, the difficulty of assignment becomes unfavorable when employing only a set of standard experiments such as 3D HNCA, 3D HN(CO)CA, 3D HNCB, 3D HN(CA)CO, 3D HNCACB, and 3D CBCA(CO)NH, which are routinely used for backbone assignment of folded biological molecules.²³ Problems of spin system overlap and degeneracy can be partly overcome by the incorporation of sequence-dependent amide nitrogen ^{15}N chemical shifts. This was exploited by applying HNN and HN(C)N triple-resonance experiments that sequentially correlate ^1HN and ^{15}N resonance chemical shifts.²⁴ In total, 181 of 207 non-proline ^1HN and ^{15}N backbone chemical shifts were assigned (87.4%). $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts were allocated for 195 residues (89.0%). For the N-terminal stretch containing residues His46-His49, no signal could be found presumably because of conformational dynamics as indicated by the very low signal intensity of residue Val50 in the HSQC spectra. Only the two initial residues of a polyaspartic stretch from Asp99 to Asp105 could be unambiguously assigned. The 17 remaining unassigned residues included Gln53 His57, Asn63, Asp64, Glu77, Asp89, Val90,

Glu116 Gln178, His199, Arg207, Arg213, Asn260, and Glu261, which could not be allocated unambiguously because of intense spin system overlap in both ^1HN and ^{15}N dimensions. The assignments have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu>) as BMRB entry 15519.12

Synchrotron Radiation Circular Dichroism Analysis and Secondary Structure Content of OPN

Synchrotron radiation CD (SRCD) spectroscopy was used to investigate the local secondary structure content of human OPN in solution. Because of the high intensity of the synchrotron ring light flux, SRCD spectroscopy allows in addition the acquisition of data in the vacuum ultraviolet wavelength region below 190 nm. Therefore, secondary structures (especially unordered structures) can be better distinguished using SRCD, as compared to conventional CD spectroscopy. In addition, the greater light intensity of the synchrotron ring gives an improved signal-to-noise ratio, thereby considerably reducing the amount of protein required and allowing measurements in buffers with higher salt contents.³⁷

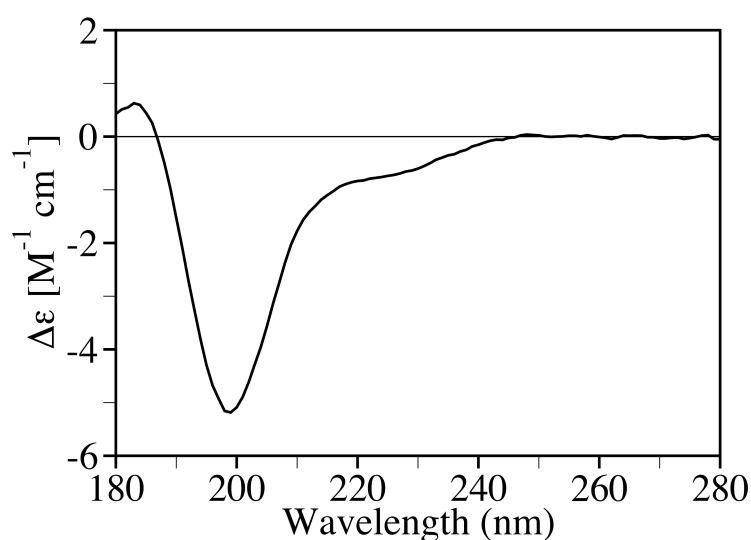


Figure 2. Synchrotron radiation circular dichroism (SRCD) spectra of OPN in 50mM NaCl and 50mM sodium phosphate (pH 6.5). Deconvolution analyses indicate the following secondary structure components are present: 5% α -helix, 13% β -sheet, and 82% other or unordered.

The SRCD spectrum of human OPN is shown in Figure 2. The spectrum is typical for a polypeptide lacking substantial amounts of regular secondary structures (i.e., helix and sheet) and has the appearance of a spectrum arising from a mostly disordered protein. Indeed, deconvolution analyses based on the SRCD data indicate the protein has a secondary structure content of 5% α -helices, 13% β -sheet, and 82% “other” or “unordered” structure (which encompasses all nonregular secondary structure types). These findings are consistent with RDC and secondary chemical shift data suggesting α -helical propensities in the region of residues 58–68 (which would also correspond to 5% α -helix) and β -sheet propensities for regions surrounding Asp80 and Ala125 (see below). The large “other” component supports the idea that OPN has a largely unstructured protein conformation.

Secondary Chemical Shifts

The differences between the observed chemical shifts of nucleus spin types and their random coil values assigned in unfolded conformations serve as a valuable tool for evaluating the residual preferences for protein secondary structure. Figure 3 shows the sequence-corrected secondary $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts for OPN.

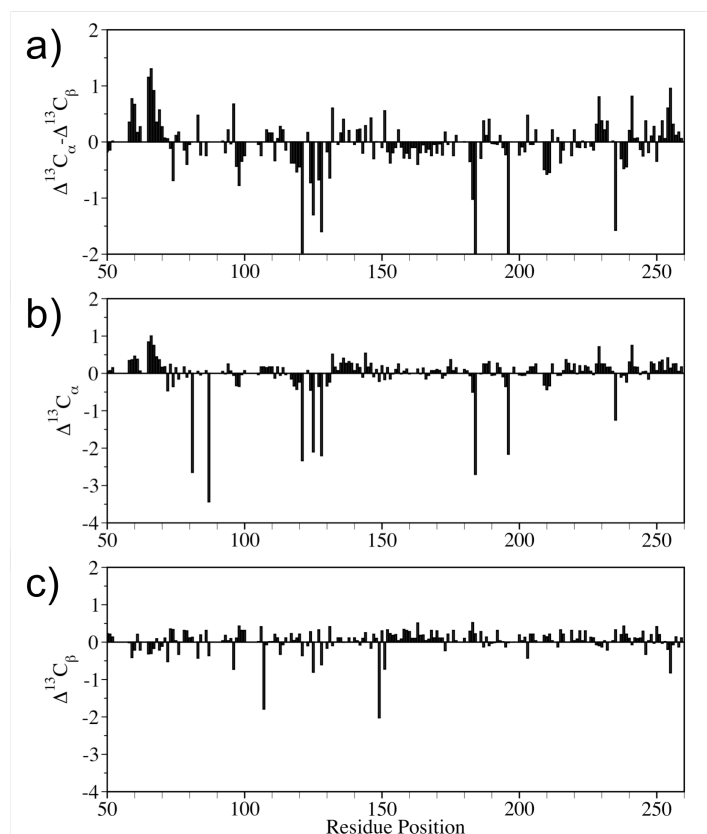


Figure 3. Residue plots of secondary ($^{13}\text{C}\alpha$ - $^{13}\text{C}\beta$), $^{13}\text{C}\alpha$, and $^{13}\text{C}\beta$ chemical shifts. Shown are $\Delta^{13}\text{C}\alpha - \Delta^{13}\text{C}\beta$ (a), $\Delta^{13}\text{C}\alpha$ (b), and $\Delta^{13}\text{C}\beta$ (c) chemical shift deviations from random coil values in OPN. Random coil values were derived from ref 38. Significant α -helical propensities were found for residues 60-70 and 130-150, whereas residues 120-130 display extended (β -strand-like) conformations. Additionally, small residue stretches in the C-terminus of OPN also display non-zero α -helical probabilities.

In α -helices, the $^{13}\text{C}\alpha$ values appear to be shifted downward relative to their random coil values, whereas the $^{13}\text{C}\beta$ values are shifted upward, leading to opposite signed effects of both nuclear spins. In β -strands, the reverse applies. Overall, the majority of the residues display chemical shift values that are typical for random coils. However, for some residues, small secondary shifts indicating a slight tendency toward a α -helical conformation are apparent, like Leu58 to Leu68, or the regions around Ala150 and His255. Slight residual preferences for a β -strand could be inferred from the chemical shifts for regions around residues Asp80 and Ala125. Additionally, we also used the TALOS database system to probe the existence of local secondary structure elements.³⁹ Overall, only a small

number of residues converged to unique solutions in Ramachandran space, thus underscoring the finding that OPN displays significant conformational flexibility. Pronounced tendencies to adopt (non-random coil) β -strand conformations were found for only a small number of residues: Thr71, Thr97, Thr119, Ser141, the fragment of residues Val152–Lys154, Lys157, and Ile158.

NMR Relaxation Rates and Solution Motional Behavior of OPN

Measurement of NMR relaxation rates provides a window on protein dynamics over a broad range of time scales.²³ The longitudinal ^{15}N -T₁ values, transverse ^{15}N -T₂ values, and heteronuclear steady-state NOE ^{15}N - $\{^1\text{H}\}$ attenuation factors are sensitive to dynamics on the picosecond to nanosecond time scale, whereas ^{15}N -T₂ can also be affected by conformational or chemical exchange processes on the millisecond to microsecond time scale. Despite the rather small chemical shift dispersion of OPN in the ^1H - ^{15}N HMQC spectrum, ^{15}N relaxation data could be obtained for 122 of 183 assigned non-proline residues at 298 K and 600 MHz. As indicated in Figure 4a, ^{15}N -T₁ values exhibit a strictly monotonous and featureless distribution along the amino acid sequence except at the C-terminal end, with an average value of 619 ms. The residue plot of the experimental ^{15}N -T₂ values (Figure 4b) shows several sequential variations ranging from 114 to 747 ms. This is in strong contrast to an expected parabolic form with a ^{15}N -T₂ minimum in the middle of the sequence arising from a complete random structure without any structural preferences.

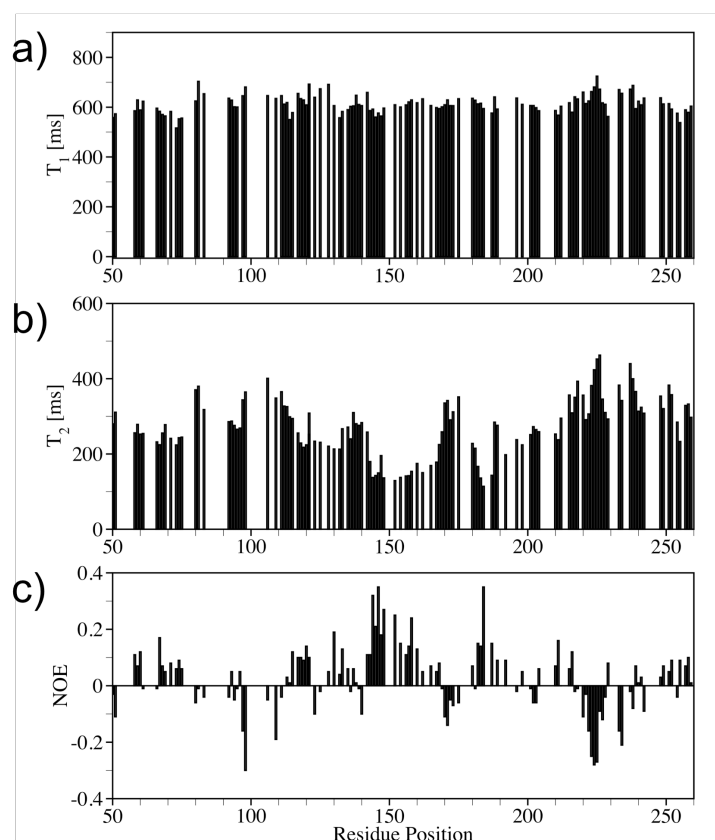


Figure 4. ^{15}N relaxation data of OPN measured at 600 MHz. ^{15}N -T₁ (a), ^{15}N -T₂ (b), and steady-state ^{15}N - $\{^1\text{H}\}$ heteronuclear NOE (c) plotted as a function of residue position.

Overall, there is good agreement between secondary chemical shifts indicating secondary structure formation and motional restrictions probed by ^{15}N -T₂ relaxation. Interestingly, the Leu182-Trp183-Trp184 stretch also appears to be constrained presumably because of the bulky and sterically demanding indole side chains. These findings are in good agreement with the obtained steady-state NOE ^{15}N - $\{^1\text{H}\}$ attenuation factors (Figure 4c) that are found to be generally higher and strictly positive in these ranges. The highest values (0.2-0.37) were observed in the middle of the sequence and for residue Trp184 (0.35). Extensive patches of negative attenuation factors were apparent from residue Ala170 to Gly175 and more pronounced along a nine-residue stretch comprising residues Gly217 Gly225, a motif consisting of residues of small side chains (three glycines, two alanines, and two serines). In addition, these parts display higher ^{15}N -T₂ values of 350-450 ms. The total average of ^{15}N -T₂ is 280

ms. The intramolecular dynamics of OPN on a microsecond to millisecond time scale was investigated by state-of-the-art ^{15}N CPMG dispersion experiments. However, analysis of the effective relaxation rate R_2 versus the CPMG frequency (ν_{CPMG}) profile revealed flat dispersion profiles for all residues resolved in the 2D ^1H - ^{15}N HSQC spectra indicating a lack of microsecond to millisecond time scale motions in OPN.

Mapping of the Heparin Interaction Site

It has been reported that OPN binds to heparin, although no clear experimental evidence was given.^{40,41} To provide more specific information about the binding site and the mode of binding, we applied ^1H - ^{15}N HSQC spectroscopy. Figure 5a shows an overlay of HSQC spectra of unligated OPN and after binding to heparin. The chemical shift differences between the heparin-free and -bound forms of OPN220 plotted as a function of residue position are shown in Figure 5b.

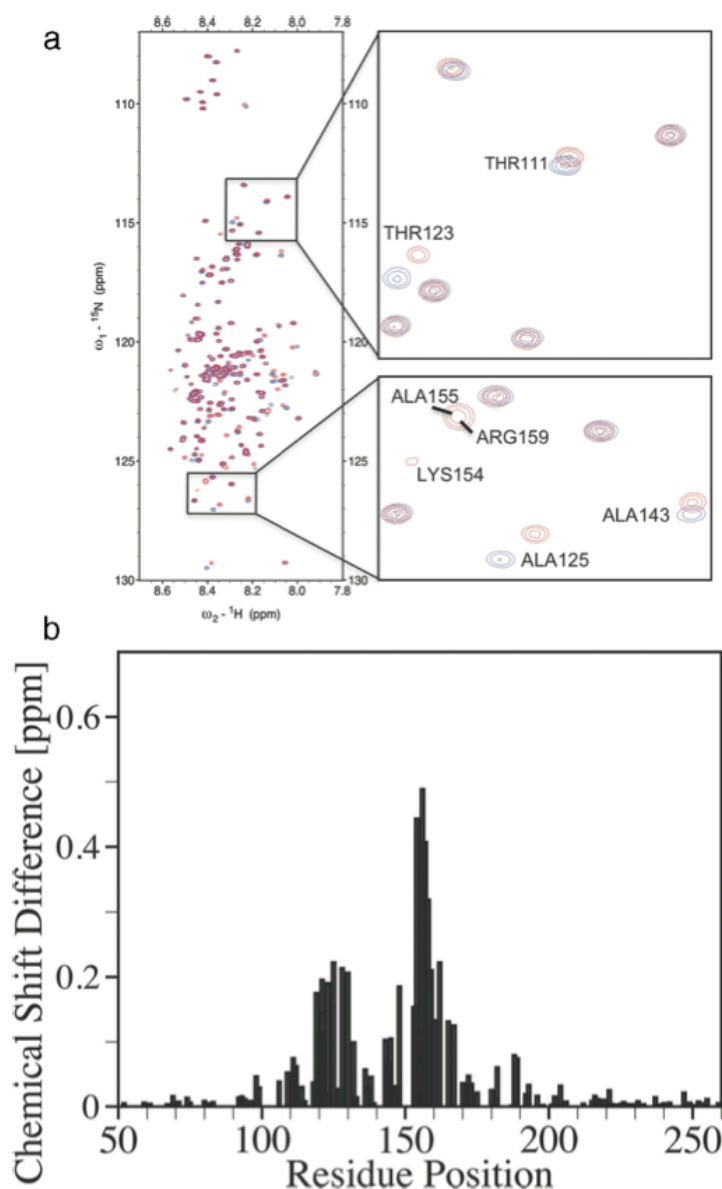


Figure 5. Heparin binding site of OPN. (a) Mapping of the heparin binding site was carried out by monitoring ^{15}N - ^1H chemical shift changes of OPN via 2D ^1H - ^{15}N HSQC spectroscopy. (b) Chemical shift differences between the heparin-free and -bound form of OPN220 plotted as a function of residue position. Measurements were performed at 600 MHz with a final heparin concentration of 1 mM.

There is evidence of a specific interaction between OPN and heparin. Although the majority of peaks were unchanged, several residues experienced significant chemical shift changes upon heparin binding. Overall, the magnitudes of chemical shift changes are small, presumably because heparin lacks aromatic ring systems and because ^1H - ^{15}N HSQC spectra report on the protein backbone

and not on side chains, which are in direct contact with the ligand. Mapping of the residues displaying frequency changes upon binding of heparin on the protein backbone clearly reveals a distinct binding site near the central integrin-binding domain (residues 117-130 and 142-167). Furthermore, OPN is an inhibitor of hydroxyapatite (HA)⁴²⁻⁴⁴ and calcium oxalate (CaOx) crystallization.⁴⁵ To elucidate the effects of Ca^{2+} on OPN, we performed ^1H - ^{15}N HSQC spectroscopy. The ^1H - ^{15}N HSQC spectra remained nearly unchanged even at a 40:1 molar excess of calcium ions over OPN (data not shown), thus further corroborating the previous results showing that post-translational modification (phosphorylation) is an important prerequisite for effective Ca^{2+} binding of OPN.^{41,46}

Meta-Structure Analysis

The meta-structure concept was recently introduced as a novel theoretical framework for protein sequence analysis.¹⁴ The protein meta-structure information provides two quantitative parameters, compactness and local secondary structure, describing the topology of a protein. The local residue-specific compactness value (C_i) is a quantitative parameter describing the structural complexity of an individual residue in the context of the 3D protein fold.^{14,47} Residues that are on average deeply buried in the interior of a structure exhibit large C_i values, whereas small C_i values are characteristic of flexible loop regions and/or intrinsically unfolded segments of the polypeptide chain. The predicted local secondary structure parameter S_i referring to the local secondary structure is defined in analogy to the well-established $^{13}\text{C}\alpha$ secondary shifts. Positive S_i values are indicative of α -helix and negative S_i values the presence of an extended conformation. For the sequence-based prediction of protein disorder, several bioinformatic approaches that are primarily based on neural networks have been developed.⁴⁸⁻⁵⁴ Meta-structure analysis differs from these bioinformatics approaches by providing quantitative information defined on a per residue basis, which can be used to analyze details of protein folds on a residue resolution level. To evaluate the performance and usefulness of the meta-structure approach for the characterization of IDPs, we compared our data to disorder prediction results obtained with DISOPRED2.⁵³ The sequence-

derived compactness data for OPN and the comparison with DISOPRED2 results as well as steady-state $^{15}\text{N}\{-^1\text{H}\}$ heteronuclear NOE are shown in Figure 6. For the calculation of compactness values, the 264-amino acid full-length quail OPN was used. For OPN, the average residue compactness (ARC) value of ~ 182 was observed, which is significantly lower than the average protein database (PDB) value of 300, suggesting that OPN does not form a stable tertiary fold in solution. Figure 6 shows a comparison of DISOPRED2 results (a) with predicted local compactness values (Figure 6b) and steady-state $^{15}\text{N}\{-^1\text{H}\}$ heteronuclear NOEs (Figure 6c).

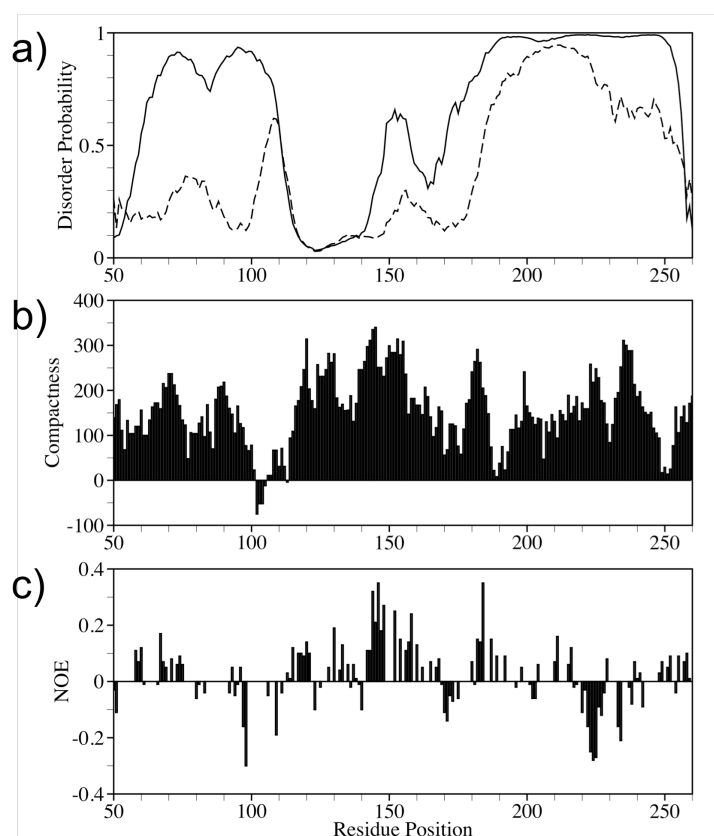


Figure 6. Meta-structure analysis and comparison with DISOPRED2 and NMR relaxation data. DISOPRED2-derived disorder propensities (a) and meta-structural parameter compactness (b) and steady-state $^{15}\text{N}\{-^1\text{H}\}$ heteronuclear NOE (c) plotted as a function of residue position. Large compactness values indicate residue positions with a tendency to form stabilizing residue-residue contacts, whereas small values are found for residues exposed to the solvent and devoid of stabilizing inter-residue interactions. Note that for the calculation full-length OPN (264 amino acids) was used, whereas the recombinant OPN applied for NMR analysis consists of 220 amino acids. The compactness values of the

truncated version were nearly unchanged. DISOPRED2 was used to calculate the disorder propensities for both full-length OPN (—) and the truncated version (---).

The predicted local secondary structural features determined by meta-structure analysis (Figure 7) correlate reasonably well with the NMR-based identification of secondary structures. The locations of the N-terminal α -helix (Gln59-Ser74) and of the two extended regions (Glu77-Phe87 and Val117-Arg132) were correctly reproduced by the meta-structure parameters. For comparison, the primary sequence of OPN was also subjected to the well-established disorder prediction server DISOPRED2.⁵³ The obtained prediction results indicate that OPN exists as an intrinsically unstructured protein in solution. Specifically, in the C-terminal region of OPN beyond residue position 190, disorder probability values larger than 0.6 were obtained, whereas in the N-terminus, disorder probabilities were consistently smaller, ranging from 0.005 at residue 125 to 0.42 at residue 28. For the most compact segment of residues 115-190, the program consistently provides minimum disorder probabilities of <0.1. The DISOPRED2 secondary structure prediction expressed in psipred values also provides evidence of the presence of small local secondary structure elements in OPN. While the α -helical psipred values identified for residues 165-170 are in good agreement with both our meta-structure prediction and NMR secondary shift data, the helical segment of residues 55-70 both predicted by meta-structure analysis and experimentally observed by NMR was not identified by DISOPRED2. β -Sheets were predicted by DISOPRED2 for residues 119-121 and 143-148. For residues 119-127, the existence of a β -sheet was also indicated by meta-structure analysis and NMR spectroscopy. In contrast, the stretch of residues 143-148 did not display significantly extended conformations. Overall, meta-structure analysis provided reliable information about the structural preferences of OPN in solution and about the presence and location of secondary structure elements that nicely correlate with experimental NMR data. Interestingly, residues displaying larger compactness values (centered around residues 125, 150, and 180) consistently exhibited reduced local backbone mobility as evidenced by ^{15}N relaxation data. This observation underscores the reliability of the predicted compactness values and suggests fruitful applications

to experimental NMR spectroscopy of IUPs (identification of spin-label attachment sites) (see below).

Residual Dipolar Couplings and Paramagnetic Relaxation Enhancements

Recently, it has been shown that residual dipolar couplings (RDCs) and paramagnetic relaxation enhancement data (PRE) are versatile spectral probes for the characterization of structural dynamics of intrinsically or natively unfolded proteins.⁵⁵⁻⁵⁷ RDCs report on time- and ensemble-averaged conformations of polypeptide chains and have been successfully used to study unfolded proteins in solution.^{56,58,59} The partial alignment of molecules allows the detection of small degrees of backbone compaction. Measurements of ^1H - ^{15}N backbone RDCs in aligned *pf*-1 phage media (10 mg/mL) probe all accessible vector bond orientations up to tens of microseconds.

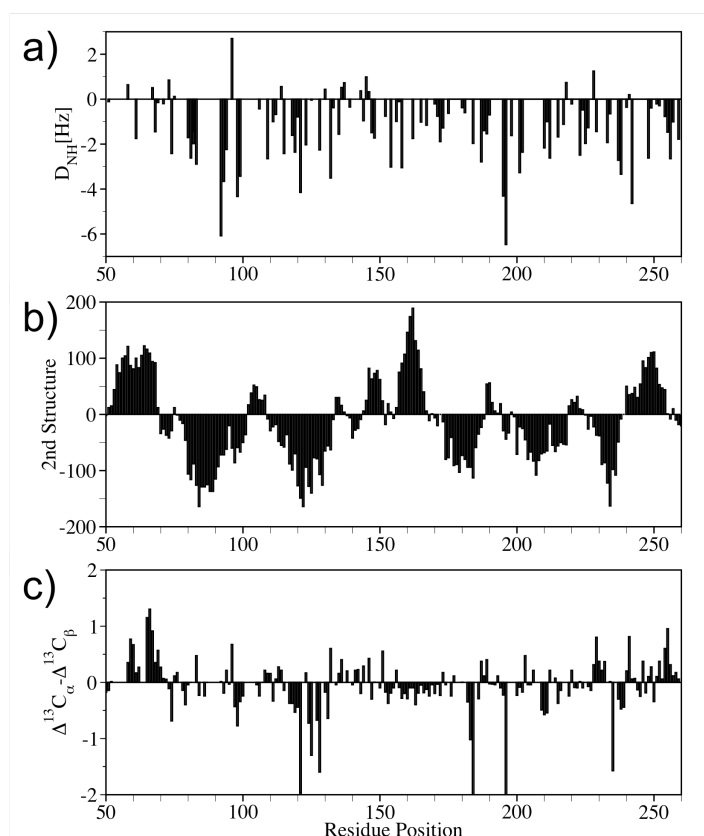


Figure 7. Secondary structure of OPN. Comparison between ^1H - ^{15}N residual dipolar couplings (RDCs) (a), meta-structure-derived secondary structure (b), and $^{13}\text{C}\alpha$ - $^{13}\text{C}\beta$ NMR secondary shifts (c). Meta-structure-derived local secondary

structure values are defined in a similar manner compared to NMR $^{13}\text{C}\alpha$ secondary shifts. Positive values are indicative of α -helical regions, whereas continuously negative values are typical for extended or β -strand regions. OPN was aligned in a buffer containing 50 mM NaCl and 50 mM sodium phosphate (pH 6.5) supplemented with a colloidal *pf*-1 phage suspension (10 mg/mL). Note that the RDCs are plotted according to convention on an inverted axis where negative values are shown as positive ones.

Figure 7 shows experimental RDC values as a function of residue position. Overall RDC values were found to be rather small. The RDC profile of OPN significantly deviates from a distribution expected for an unfolded polypeptide chain represented by a random flight or Gaussian model for which non-vanishing RDCs displaying rather uniform sequence position dependencies can be expected. Consistently, greater-amplitude RDCs were found for residues displaying extended conformations indicated by both negative secondary NMR chemical shifts and secondary meta-structure values. Changes in the signs of RDCs were observed for residues embedded in α -helical segments in the N-terminal half of the 220-amino acid OPN protein at positions 50-70 and 130-160. As a result, the RDC data indicate that the solution structure of OPN is best characterized as a flexible polypeptide chain with distinct local structure elements.

PREs report on long-range interactions and thus provide valuable information about transient contact formation in unfolded proteins.^{57,59,60} Despite the ease of experimental implementation, one of the inherent problems is the choice of suitable spin-label sites, e.g., sites of Cys mutations. Here we have used meta-structure-derived compactness data to identify suitable sites of spin-label attachment. Figure 8a shows the compactness as a function of residue position. The sites of spin-label attachment were selected on the basis of minimal compactness value, as for these regions tight side chain interactions or packing can presumably be neglected. Panels b and c of Figure 8 show PRE data obtained for OPN mutants S108C and S188C, respectively.

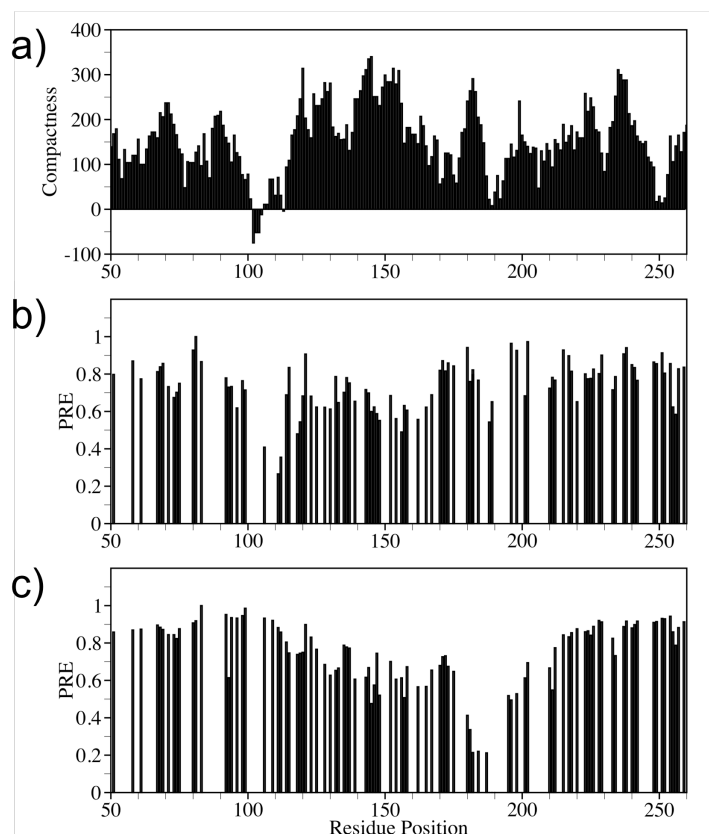


Figure 8. Residue plot of protein compactness and paramagnetic relaxation enhancement (PRE) for OPN. (a) The protein compactness data were used to identify possible sites for spin-label attachment. Residues displaying small compactness values are found in exposed and/or conformationally flexible parts of the protein and thus most likely unperturbed by the attachment of a paramagnetic spin-label. (b and c) PRE data obtained for Cys108 (b) and Cys188 (c). Significant PREs with smaller intensity ratios were found for the stretch of residues 110-165. This is in good agreement with larger compactness values predicted for this part of the polypeptide chain shown in panel a.

It can be seen that ^{15}N paramagnetic relaxation enhancements are not limited to only neighboring residues but also appear in regions distant from the affected mutation sites. This indicates some degree of compaction and transient contact formation. Most importantly, residues with significant PREs are observed in regions for which both positive heteronuclear NOEs were observed and large compactness values were predicted like residues 115-160. The pulse field gradient (PFG) NMR diffusion data and the large hydrodynamic radius (see below) indicate that OPN exists in solution as a largely extended polypeptide chain with some local compaction in the central domain that encompasses the

integrin-binding domain. Independent NMR evidence corroborating this collapsed state was recently provided using paramagnetic relaxation enhancement data where transient interactions between the N- and C-terminal domains were observed.¹³

Measurement of Hydrodynamic Radii

We determined by PFG NMR diffusion that the hydrodynamic radius R_h of OPN is 41.07 Å. According to the empirical relationship between the length of a polypeptide chain and its hydrodynamic radius,³¹ a globular 220-residue protein should have an R_h of 22.7 Å, whereas the same unfolded peptide chain should exhibit an R_h of 47.8 Å. The R_h value of OPN is significantly larger than the expected value for a globular protein, indicating that under these conditions OPN is predominantly unfolded. Nevertheless, the lower R_h of OPN compared to the expected value for a completely unfolded chain suggests that OPN retains some compact structures of minimal enthalpy stabilized by favorable side chain interactions.

Independent experimental data for the solution behavior of OPN were provided by a small-angle X-ray scattering analysis. The one-dimensional scattering curves were further treated using GIFT.³⁴⁻³⁶ Because of the low protein concentration, the contribution of protein interactions is negligible. Furthermore, inter-molecular electrostatic interactions caused by ions in the solution buffer also do not contribute to the scattering curve. The scattering was therefore only related to the overall shape, the electron density, and its distribution within the protein. The pair distance distribution function (PDDF) could be obtained by IFT. This function can be visualized as a histogram of distances present in the protein. The shape of the PDDF can be evaluated to gain information about the structure of the protein. For a prototypical Gaussian coil polymer, the PDDF shows a linear increase at short distances and a slow decay at longer distances. Figure 9 shows the PDDF extracted from the experimental scattering curve of OPN.

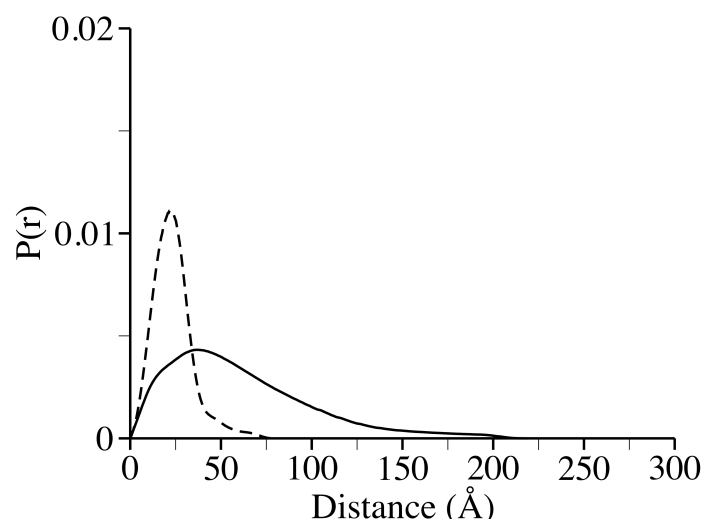


Figure9. Distance distribution function $P(r)$ extracted from SAXS data using a generalized inverse Fourier transform technique (GIF).^{35,36} The ensemble-averaged distance distribution function of osteopontin is shown as a solid line. For comparison, the distance distribution function of a stably folded protein (quail Q83) with a rather compact 3D structure is also shown (---).

While the linear increase could be clearly observed in the PDDF of the protein, the decrease in the PDDF function is flatter than that one would expect for a Gaussian coil. The maximum dimension of the protein could be estimated to $\sim 210 \text{ \AA}$. Furthermore, a shoulder in the function at $\sim 15 \text{ \AA}$ was observed. It is important to mention that the scattering curve represents a sum of the scattering, which is caused by all structures appearing in the measured volume during the measurement time. The extracted PDDF clearly indicates that OPN comprises a heterogeneous conformational space containing both extended coil and compact conformation with distinct side chain interactions. For comparison, Figure 9 also shows the PDDF curve of the stably folded lipocalin protein Q83⁶¹ that strikingly differs from that of OPN. The average gyration radius of OPN was determined to be $51.3 \text{ \AA}$. For comparison, the average gyration radius of Q83 was determined to be $20.7 \text{ \AA}$, which is in very good agreement with the 3D structural data obtained by NMR spectroscopy.⁶¹ The data obtained for OPN suggest that the analysis of PDDFs provides valuable information about the conformational ensembles of intrinsically disordered proteins in solution. Experiments exploiting these features are currently underway in our laboratories.

DISCUSSION

The solution structural features and local dynamics of OPN were investigated by a combined biophysical approach involving meta-structure-based sequence analysis, multidimensional NMR spectroscopy, NMR translational diffusion measurements and ^{15}N relaxation parameters, SRCD spectroscopy, and small-angle X-ray scattering (SAXS). All biophysical data consistently demonstrated that OPN exists as a largely disordered protein in solution and thus constitutes another member of the growing intrinsically disordered protein family. These results are in agreement with a recent report identifying OPN as an intrinsically disordered protein.¹³ However, these broad-based and in-depth structural analyses reveal that although OPN is largely unfolded in solution, significant ^{13}C secondary shifts that differ from random coil values indicate that the local secondary structure elements are only transiently formed, lacking the cooperative behavior typical for globular and stably folded proteins. The variation of local mobility along the backbone significantly differs from the profile expected for a random coil protein and is similar to profiles of other intrinsically unfolded or natively unstructured proteins undergoing preferential conformational sampling. In unfolded proteins, residues in the middle of the polypeptide chain are expected to be less mobile compared to terminal segments, leading to a bell-shaped distribution.^{62,63} The correlation between deviations from random coil chemical shift values and backbone dynamics provides convincing evidence of the existence of local structure elements involving residues that are located in reasonably well-ordered, hydrogen-bonded, and motionally restricted parts of the polypeptide chain. Local motional restriction implies that residues located in these structurally preformed segments do not populate unfolded (coil-like) conformations but rather exist as cooperatively folded local structural elements (β -strands and α -helices). Interestingly, these preformed and motionally restricted structural motifs comprise the biochemically established protein interaction motifs of OPN. The structural propensities and the overall fold of OPN were also investigated using the recently introduced meta-structure approach for sequence-based calculation of residue-specific compactness and local secondary structural

features in proteins.¹⁴ The predicted average residue compactness values of OPN amount to ~ 182 , which is clearly below values observed for stably folded proteins and thus indicative of an intrinsically unstructured protein. However, analyzing the compactness values on a per residue basis provides more insight into the meta-structure characteristics and shows that OPN consists of several segments with some residual local structures. Overall, the local compactness values reasonably agree with NMR-derived local structural ordering. For example, residues located in the central subdomain extending from residue 110 to 190 and comprising several integrin-binding sites display larger than average compactness values. Most importantly, relatively large and positive NOE values were observed for the majority of residues located in this stretch of amino acids exhibiting larger than average compactness values.

The meta-structure approach also proved to be valuable, allowing a rational design of PRE experiments performed in this study. While PRE is well-established as a powerful NMR tool for the identification of transient long-range contacts in disordered proteins, the general applicability is somehow hindered by the selection of residue positions for the placement of spin-labels. As we have shown in this study, the meta-structure-derived compactness values are useful for identifying suitable attachment sites exclusively on the basis of the primary sequence. On the basis of our findings, we propose to use minimum compactness values as the selection criterion for the identification of spin-label sites. The obtained PRE results confirm the heterogeneous flexibility along the OPN backbone. Consistently, significant PRE values are found for residues located in regions with larger than average compactness values indicating the transient long-range contact probabilities of these segments.

The use of SRCD, as opposed to conventional CD spectroscopy, to study the secondary structure of the protein was important in this study for several reasons:³⁷ to allow the accurate evaluation of the “other” or disordered components, which produce signals in the low-wavelength region below 200 nm, to allow measurements under high-salt conditions, and to provide high-quality, good signal-to-noise levels in samples while using minimal amounts of material. Also, because of the agreement with the results obtained with the other techniques in this study, this protein now provides an excellent example of an

unfolded protein that can be used in reference data sets for subsequent analyses of other proteins.

We also investigated the binding of OPN to the ligand heparin. The heparin binding site was mapped by ^1H - ^{15}N HSQC to residues 117-167. Interestingly, these regions display less flexibility and exist in locally restrained conformations. This observation suggests that the more compact regions in OPN facilitate the accommodation of relevant binding partners (e.g., heparin and integrin receptor binding site) proceeding via conformational selection mechanisms, presumably because of the reduction of the entropic penalty in the Gibbs free energy of binding.

From the data presented here, a model in which transient secondary structure and tertiary structure coexist and establish a diverse and heterogeneous conformational ensemble emerges. The individual transiently formed structures of the ensemble are largely different, ranging from extended coils devoid of stabilizing interactions to compact structures with accentuated electrostatic surfaces and exposed binding sites, preformed to accommodate authentic binding partners of OPN. The flexibility of the polypeptide chain partly uncouples binding events at the individual interaction sites lined up along the OPN backbone. This modular architecture and intrinsic flexibility may thus be a necessary prerequisite for the biological functionality of OPN as it allows for substantial structural plasticity, necessary to accommodate various protein binding partners recruited to the cellular surface.

FUNDING

This work was supported by Grants P17041 (to K.B.) and P18148 (to M.H.) from the Austrian Science Foundation (FWF) and W 1221-B03, EU-BACRNA, SFB-17, and WWTF LS612 from the FWF (all to R.K.). R.A. is a recipient of a DOC-ffORTE fellowship of the Austrian Academy of Sciences. N.C. is a recipient of a Lise-Meitner FWF fellowship. The SRCD work was supported by project grants from the U.K. Biotechnology and Biological Sciences Research Council (BBSRC) to B.A.W. and by synchrotron beam time grants to B.A.W. Access to the CD1 beam-line at ISA is acknowledged under the EU Integrated

Infrastructure Initiative (I3), Integrated Activity on Synchrotron and Free Electron Laser Science (IA-SFS), Contract RII3-CT-2004-506008.

ACKNOWLEDGMENTS

We thank Renate Rieder and Kathrin Breuker (Institute of Organic Chemistry, University of Innsbruck) for performing ESI mass spectrometry

REFERENCES

1. Bellahcene, A.; Castronovo, V.; Ogbureke, K. U.; Fisher, L. W.; Fedarko, N. S., Small integrin-binding ligand N-linked glycoproteins (SIBLINGS): multifunctional proteins in cancer. *Nat Rev Cancer* **2008**, *8* (3), 212-26.
2. Weber, G. F., The metastasis gene osteopontin: a candidate target for cancer therapy. *Biochim Biophys Acta* **2001**, *1552* (2), 61-85.
3. Rittling, S. R.; Chambers, A. F., Role of osteopontin in tumour progression. *Br J Cancer* **2004**, *90* (10), 1877-81.
4. Wai, P. Y.; Kuo, P. C., Osteopontin: regulation in tumor metastasis. *Cancer Metastasis Rev* **2008**, *27* (1), 103-18.
5. Hartl, M.; Karagiannidis, A. I.; Bister, K., Cooperative cell transformation by Myc/Mil(Raf) involves induction of AP-1 and activation of genes implicated in cell motility and metastasis. *Oncogene* **2006**, *25* (29), 4043-55.
6. Gardner, H. A.; Berse, B.; Senger, D. R., Specific reduction in osteopontin synthesis by antisense RNA inhibits the tumorigenicity of transformed Rat1 fibroblasts. *Oncogene* **1994**, *9* (8), 2321-6.
7. Rangaswami, H.; Bulbule, A.; Kundu, G. C., Osteopontin: role in cell signaling and cancer progression. *Trends Cell Biol* **2006**, *16* (2), 79-87.
8. Denhardt, D. T.; Noda, M.; O'Regan, A. W.; Pavlin, D.; Berman, J. S., Osteopontin as a means to cope with environmental insults: regulation of inflammation, tissue remodeling, and cell survival. *J Clin Invest* **2001**, *107* (9), 1055-61.
9. Junaid, A.; Amara, F. M., Osteopontin: correlation with interstitial fibrosis in human diabetic kidney and PI3-kinase-mediated enhancement of expression by glucose in human proximal tubular epithelial cells. *Histopathology* **2004**, *44* (2), 136-46.
10. Okamoto, H., Osteopontin and cardiovascular system. *Mol Cell Biochem* **2007**, *300* (1-2), 1-7.
11. Prince, C. W.; Oosawa, T.; Butler, W. T.; Tomana, M.; Bhowan, A. S.; Bhowan, M.; Schrohenloher, R. E., Isolation, characterization, and biosynthesis of a phosphorylated glycoprotein from rat bone. *J Biol Chem* **1987**, *262* (6), 2900-7.
12. Schedlbauer, A.; Ozdowy, P.; Kontaxis, G.; Hartl, M.; Bister, K.; Konrat, R., Backbone assignment of osteopontin, a cytokine and cell attachment protein implicated in tumorigenesis. *Biomol NMR Assign* **2008**, *2* (1), 29-31.
13. Yamaguchi, Y.; Hanashima, S.; Yagi, H.; Takahashi, Y.; Sasakawa, H.; Kurimoto, E.; Iguchi, T.; Kon, S.; Uede, T.; Kato, K., NMR characterization of intramolecular interaction of osteopontin, an intrinsically disordered protein with cryptic integrin-binding motifs. *Biochem Biophys Res Commun* **2010**, *393* (3), 487-91.
14. Konrat, R., The protein meta-structure: a novel concept for chemical and molecular biology. *Cell Mol Life Sci* **2009**, *66* (22), 3625-39.
15. Lees, J. G.; Smith, B. R.; Wien, F.; Miles, A. J.; Wallace, B. A., CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal Biochem* **2004**, *332* (2), 285-9.
16. Miles, A. J.; Wien, F.; Lees, J. G.; Rodger, A.; Janes, R. W.; Wallace, B. A., Calibration and standardisation of synchrotron radiation circular dichroism and

conventional circular dichroism spectrophotometers. *Spectroscopy-an International Journal* **2003**, 17 (4), 653-661.

17. Whitmore, L.; Wallace, B. A., DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **2004**, 32 (Web Server issue), W668-73.

18. Sreerama, N.; Venyaminov, S. Y.; Woody, R. W., Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem* **2000**, 287 (2), 243-51.

19. Provencher, S. W.; Glockner, J., Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **1981**, 20 (1), 33-7.

20. van Stokkum, I. H.; Spoelder, H. J.; Bloemendal, M.; van Grondelle, R.; Groen, F. C., Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem* **1990**, 191 (1), 110-8.

21. Sreerama, N.; Woody, R. W., Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* **2000**, 287 (2), 252-60.

22. Kay, L. E.; Keifer, P.; Saarinen, T., Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *Journal of the American Chemical Society* **1992**, 114 (26), 10663-10665.

23. Cavanagh, J.; Fairbrother, W.; Palmer, A. G.; Skegton, N., Protein NMR Spectroscopy. *Academic Press, New York* **2007**.

24. Panchal, S. C.; Bhavesh, N. S.; Hosur, R. V., Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and ¹⁵N sequential correlations in (¹³C, ¹⁵N) labeled proteins: application to unfolded proteins. *J Biomol NMR* **2001**, 20 (2), 135-47.

25. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995**, 6 (3), 277-93.

26. Goddard, T. D.; Kneller, D. G., Sparky 3. *University of California, San Francisco* **2002**.

27. Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D., ¹H, ¹³C and ¹⁵N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* **1995**, 5 (1), 67-81.

28. Schwarzing, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J., Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* **2001**, 123 (13), 2970-8.

29. 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.; Kay, L. E., Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **1994**, 33 (19), 5984-6003.

30. Tollinger, M.; Skrynnikov, N. R.; Mulder, F. A.; Forman-Kay, J. D.; Kay, L. E., Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc* **2001**, 123 (46), 11341-52.

31. Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J., Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999**, 38 (50), 16424-31.

32. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **1992**, 2 (6), 661-5.
33. Ottiger, M.; Delaglio, F.; Bax, A., Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **1998**, 131 (2), 373-8.
34. Orthaber, D.; Bergmann, A.; Glatter, O., SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *Journal of Applied Crystallography* **2000**, 33, 218-225.
35. Glatter, O., A new method for the evaluation of small-angle scattering data. *J. Appl. Cryst.* **1977**, 10, 415-421.
36. Glatter, O., The interpretation of real-space information from small-angle scattering experiments. *J. Appl. Cryst.* **1979**, 12, 166-175.
37. Wallace, B. A., Protein characterisation by synchrotron radiation circular dichroism spectroscopy. *Q Rev Biophys* **2009**, 42 (4), 317-70.
38. Tamiola, K.; Acar, B.; Mulder, F. A., Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* **2010**, 132 (51), 18000-3.
39. Cornilescu, G.; Delaglio, F.; Bax, A., Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **1999**, 13 (3), 289-302.
40. He, B.; Mirza, M.; Weber, G. F., An osteopontin splice variant induces anchorage independence in human breast cancer cells. *Oncogene* **2006**, 25 (15), 2192-202.
41. Kazanietki, C. C.; Uzwiak, D. J.; Denhardt, D. T., Control of osteopontin signaling and function by post-translational phosphorylation and protein folding. *J Cell Biochem* **2007**, 102 (4), 912-24.
42. Boskey, A. L.; Maresca, M.; Ullrich, W.; Doty, S. B.; Butler, W. T.; Prince, C. W., Osteopontin-hydroxyapatite interactions in vitro: inhibition of hydroxyapatite formation and growth in a gelatin-gel. *Bone Miner* **1993**, 22 (2), 147-59.
43. Goldberg, H. A.; Hunter, G. K., The inhibitory activity of osteopontin on hydroxyapatite formation in vitro. *Ann N Y Acad Sci* **1995**, 760, 305-8.
44. Hunter, G. K.; Kyle, C. L.; Goldberg, H. A., Modulation of crystal formation by bone phosphoproteins: structural specificity of the osteopontin-mediated inhibition of hydroxyapatite formation. *Biochem J* **1994**, 300 (Pt 3), 723-8.
45. Worcester, E. M.; Beshensky, A. M., Osteopontin inhibits nucleation of calcium oxalate crystals. *Ann N Y Acad Sci* **1995**, 760, 375-7.
46. Gericke, A.; Qin, C.; Spevak, L.; Fujimoto, Y.; Butler, W. T.; Sorensen, E. S.; Boskey, A. L., Importance of phosphorylation for osteopontin regulation of biomineralization. *Calcif Tissue Int* **2005**, 77 (1), 45-54.
47. Mayer, O.; Rajkowitsch, L.; Lorenz, C.; Konrat, R.; Schroeder, R., RNA chaperone activity and RNA-binding properties of the E. coli protein StpA. *Nucleic Acids Res* **2007**, 35 (4), 1257-69.
48. Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J., GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **2003**, 31 (13), 3701-8.
49. Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K., Sequence complexity of disordered protein. *Proteins* **2001**, 42 (1), 38-48.

50. Li, X.; Romero, P.; Rani, M.; Dunker, A. K.; Obradovic, Z., Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform* **1999**, *10*, 30-40.
51. Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I., IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21* (16), 3433-4.
52. Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B., Protein disorder prediction: implications for structural proteomics. *Structure* **2003**, *11* (11), 1453-9.
53. 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* **2004**, *337* (3), 635-45.
54. Zeev-Ben-Mordehai, T.; Rydberg, E. H.; Solomon, A.; Toker, L.; Auld, V. J.; Silman, I.; Botti, S.; Sussman, J. L., The intracellular domain of the Drosophila cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded. *Proteins* **2003**, *53* (3), 758-67.
55. Shortle, D.; Ackerman, M. S., Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **2001**, *293* (5529), 487-9.
56. Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M., Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J Am Chem Soc* **2005**, *127* (51), 17968-9.
57. Gillespie, J. R.; Shortle, D., Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J Mol Biol* **1997**, *268* (1), 170-84.
58. Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M., A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A* **2005**, *102* (47), 17002-7.
59. Bertoncini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M., Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc Natl Acad Sci U S A* **2005**, *102* (5), 1430-5.
60. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M., Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* **2005**, *127* (2), 476-7.
61. Hartl, M.; Matt, T.; Schuler, W.; Siemeister, G.; Kontaxis, G.; Kloiber, K.; Konrat, R.; Bister, K., Cell transformation by the v-myc oncogene abrogates c-Myc/Max-mediated suppression of a C/EBP beta-dependent lipocalin gene. *J Mol Biol* **2003**, *333* (1), 33-46.
62. Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E., Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* **1997**, *36* (9), 2390-402.
63. Farrow, N. A.; Zhang, O. W.; Szabo, A.; Torchia, D. A.; Kay, L. E., Spectral Density-Function Mapping Using N-15 Relaxation Data Exclusively. *Journal of Biomolecular Nmr* **1995**, *6* (2), 153-162.

Chapter II

Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin

Platzer G,* Kurzbach D,* Schwarz TC, Henen MA, Konrat R, Hinderberger D.
Biochemistry **2013** 52(31), 5167-5175

*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 and nuclear magnetic resonance spectroscopy, we demonstrate that the OPN ensemble exhibits not only characteristics of an extended and flexible polypeptide, as expected for an IDP, but also 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 the 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 that have stronger tendencies toward 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 to an IDP. We aim to combine data gained from a dynamic system state (NMR) with data about a static 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 the 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 by 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 for 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.²⁵] This is achieved by immersing the sample tube in supercooled isopentane. In this way, a snapshot representative of the

solution at the glass transition temperature is detected. Prior to being freeze-quenched, the samples were transferred to 3 mm outer diameter quartz tubes. The sample volume was approximately 100 μL and always large enough to fill the complete sensitive volume in the resonator. The four-pulse DEER sequence $\pi/2(\nu_{\text{obs}})-\tau_1-\pi(\nu_{\text{obs}})-(\tau_1 + t)-\pi(\nu_{\text{pump}})-(\tau_2 - t)-\pi(\nu_{\text{obs}})-\tau_2$ -echo was used to obtain dipolar time evolution data at X-band frequencies (9.2–9.4 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 μs) and τ_1 were kept constant. Proton modulation was averaged by the addition of eight time traces of variable τ_1 values, starting with a $\tau_{1,0}$ of 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:

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

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 50 mM sodium phosphate and 50 mM NaCl (pH 6.5) with 10% D₂O as the lock solvent. PRE intensity ratios were derived from pulsed field gradient (PFG) sensitivity-enhanced two-dimensional ¹⁵N–¹H HSQC spectra of ¹⁵N-labeled OPN220 mutants C54, C108, C188, and C247.²⁶ NMR spectra were processed using NMRPipe²⁷ 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-oxyl-2,2,5,5-tetramethyl- Δ 3-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 1 mM. 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 $\Delta\text{PRE} = {}^{15}\text{N}\text{--}{}^1\text{H}$ HSQC intensity (high urea and salt) – ${}^{15}\text{N}\text{--}{}^1\text{H}$ HSQC intensity (no urea or salt). The usage of two different samples in the presence and absence of free MTSL is a 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 8 M 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).

${}^{13}\text{C}\text{--}{}^1\text{H}$ HSQC NMR

Reductive methylation procedures were performed as described by Means and Feeney.²⁸ The protein was dialyzed against 10 mM HEPES, 150 mM NaCl, and 1 mM 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 ${}^{13}\text{C}$ -labeled formaldehyde were added to 0.5 mL of 0.1 mM protein, and the reaction mixture was incubated while being stirred at 4 °C. Subsequently, the addition of the borane ammonium complex and [${}^{13}\text{C}$]formaldehyde was repeated, and the reaction mixture was incubated for an additional 2 h. After 0.12 mL of a 1.6 mM borane ammonium complex solution had been added, the reaction mixture was incubated at 4 °C while being stirred 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 ${}^{13}\text{C}\text{--}{}^1\text{H}$ HSQC NMR experiments were conducted with synthesized ${}^{13}\text{C}$ -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-residue 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 one-third of the whole protein; mutants C54–C188 (2/3 S) and C108–C247 (2/3 L) each span approximately two-thirds of the protein, and mutant C54–C247 (3/3 S) spans 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 that are dependent on the urea concentration. In a second step, we investigate cooperatively compacted states by means of paramagnetic relaxation enhancements and rationalize our results on the basis of noncovalent structuring principles in OPN.

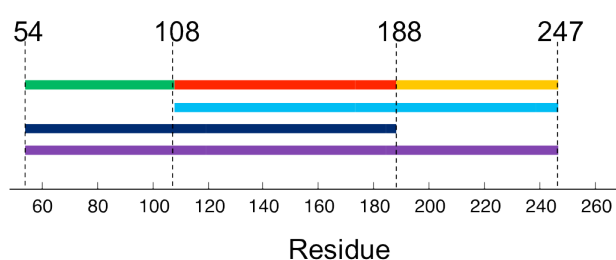


Figure 1. Scheme of spin labeling sites along the protein backbone and sketch of residues spanned by each of the six OPN double mutants.

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 of the Supporting Information for a detailed description of DEER)] that, after background correction that eliminates intermolecular contributions, display intramolecular dipole–dipole couplings between the two electrons of the spin 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

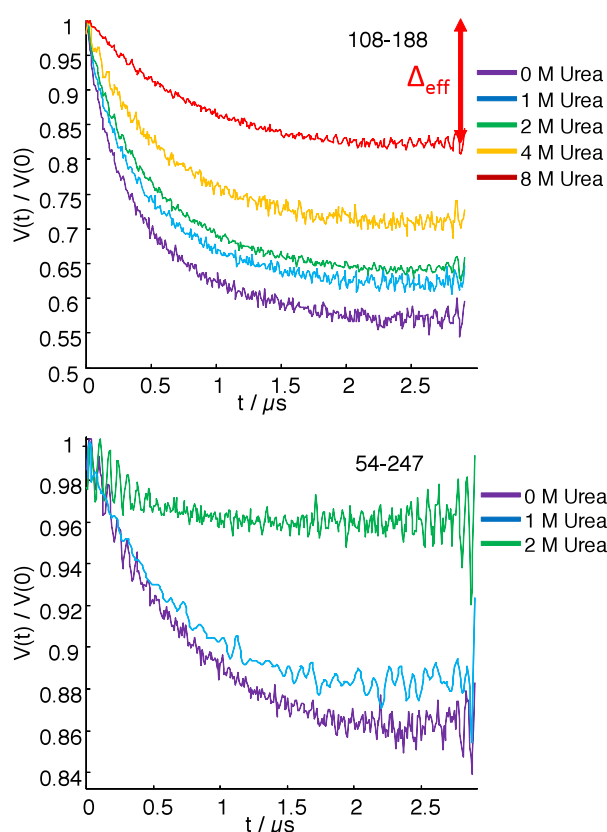


Figure 2. DEER time traces of C108–C188 (1/3 L) and C54–C247 (3/3 S) at different urea concentrations. Δ_{eff} is defined as the signal decay at 3 μs , as indicated by the double-headed arrow [note the different $V(t)/V(0)$ scales].

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 of a double mutant defining a small OPN segment, and C54–C247 (3/3 S), representative of a

large segment]. No clear-cut modulations can be observed 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 Figures S2 and S3 of the Supporting Information; all time traces are devoid of modulations).²⁹ 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 . As 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 $P(R)$ values. For broad distance distributions Δ_{eff} decreases with an increasing interspin distance R . As such, a decrease in Δ_{eff} with an 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.

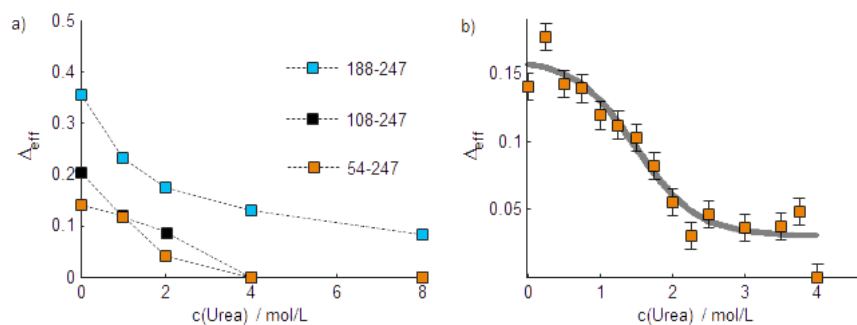


Figure 3. (a) Δ_{eff} for selected double mutants as a function of urea concentration. All the data for all double mutants under investigation can be found in Figures S2–S4 of the Supporting Information. (b) Detailed representation of Δ_{eff} for C54–C247 (3/3 S) as a function of urea concentration. 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 lengths (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 an 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 of the Supporting Information)], one observes an approximately linear decrease in 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 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 of 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).

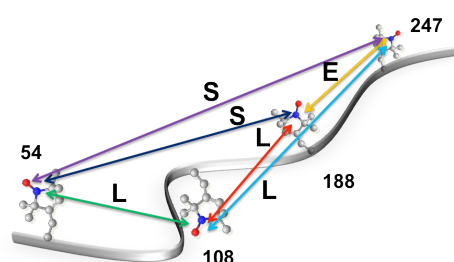


Figure 4. Sketch of OPN double mutants assessed by DEER. Labels E (exponential), L (linear), and S (sigmoidal) denote the profile shapes of the respective Δ_{eff} functions (see Figure 3 and Figures S2 and S4 of the Supporting Information).

A sigmoidal denaturation profile is indicative of 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 a 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 come closer along the primary sequence and the mean distance becomes shorter (e.g., C188–C247 (1/3 L)), longer distances more significantly 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 a significantly different urea dependence: a sigmoidal Δ_{eff} profile for mutant C54–C247 (3/3 S) and only gradual changes for double mutants spanning segments smaller 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 transition 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 transition 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 absence of 8 M urea. Figure 5a shows Δ_{eff} values for denaturation conditions with 4 M NaCl, 8 M urea, or 4 M NaCl with 8 M urea. Figure 5b shows exemplary DEER time traces for 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 an 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 has only small effects on Δ_{eff} . This can be rationalized as follows. NaCl screens electrostatic interactions, while urea does not.³⁰⁻³² 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 four corresponding 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.

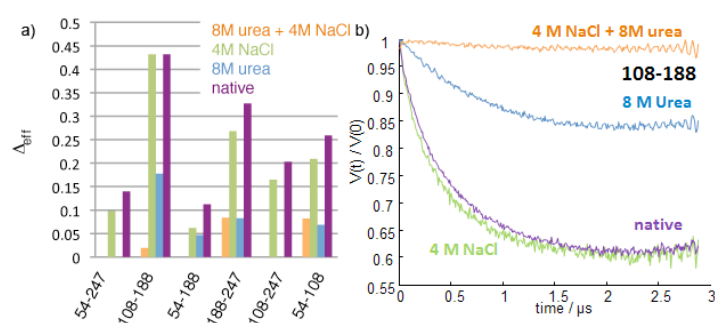


Figure 5. (a) Δ_{eff} values for the different double mutants under different denaturing conditions (4 M NaCl, 8 M urea, or 4 M NaCl with 8 M urea). (b) Exemplary (for C108–C188 (1/3 L)) DEER time traces for different denaturation conditions (4 M NaCl, 8 M urea, or 4 M NaCl with 8 M urea).

It should be noted that there is a growing body of evidence of the so-called direct mechanism of urea denaturation to describe protein–urea interaction correctly.^{33,34} This mechanism states that urea directly binds to the protein backbone likely (primarily) through dispersive interactions and thereby interrupts protein structure-stabilizing interactions.^{33,34} 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 54–108 mutants, 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.^{35,36} This discrepancy might be traced back to 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.³⁶

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 superposition of ^{15}N – ^1H NMR spectra at different urea concentration is shown in Figure S8 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. For most residues in and outside the core region, a more or less steady increase in Δcs can be observed. The core regions are approximately located between residues 100 and 180 (see Figure S5 of the Supporting Information for more cs data for residues of the core segment of OPN). 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 were 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 lower urea concentrations, cannot generally be reproduced through Δcs , the chemical shift changes clearly provide additional evidence of a more compact segment in OPN located between residues 100 and 180.¹⁶ As such, the Δcs data are not in conflict with the interpretation derived above from DEER.

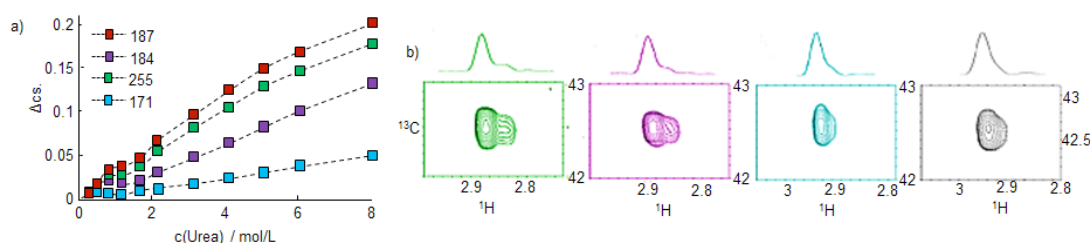


Figure 6. (a) ^{15}N - ^1H NMR chemical shift changes {calculated as $\text{cs}[c(\text{urea}) = 0 \text{ M}] - \text{cs}[c(\text{urea}) = x \text{ M}]$ } of selected backbone positions as a function of urea concentration. (b) ^{13}C - ^1H HSQC of $^{13}\text{CH}_3$ -Lys-labeled side chains: green for 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 the dependence of the urea concentration to achieve suppression of water signals.

An additional indication of the existence of compact structures in 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 a 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 a lower number of 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-based 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 sizable long-range PRE effects provides unambiguous evidence of the existence of compact states.^{37,38} 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 mutants C108 and C188 provide clear evidence of 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.

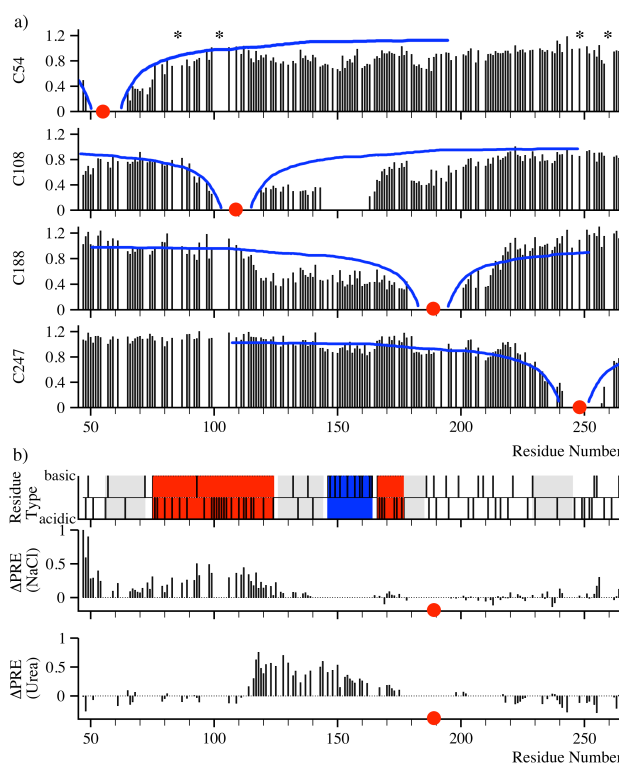


Figure 7. (a) PRE data for the four single mutants C54, C108, C188, and C247. Superimposed in blue are PREs calculated for random coils with a Flory characteristic ratio of 2 by the Solomon–Bloembergen relation.²³ The red dots mark the different labeling sites. The asterisks mark stretches comprising larger numbers of unassigned resonances. (b) 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 sizable PRE changes under high-NaCl conditions, while residues 125–150 are strongly affected by urea. Hence, from these results and the EPR 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 substates 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 residues 100 and 180 as well as a 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 the 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 ability to cooperatively sample both unfolded and cooperatively folded structures.

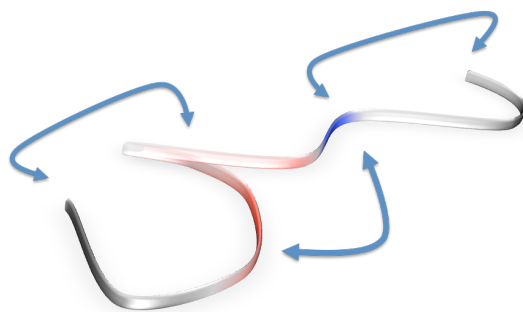


Figure 8. Sketch of the assumed “average” structure of OPN based on the PRE data. The arrows indicate significant PRE effects. As such, OPN can be pictured as having a more compact core and back-folded termini. The colors refer to the charge map in Figure 7b (blue corresponds to patches of primarily 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.³⁹ The fact that OPN samples cooperatively stabilized 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,⁴⁰ although even in the bound state IDPs (can) retain substantial conformational flexibilities (“fuzziness”),⁴¹ be it static (multiple conformations) or dynamic disorder (fluctuation between different states). Protein–protein interaction is typically described either as an induced fit or as a conformational selection fit.^{42,43} 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.⁴⁴ 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.^{40,41} 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 might thus endow OPN with unique abilities 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.⁴⁵ 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.^{33,34} 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.

FUNDING

This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG) via Grant HI 1094/2-1, the Max Planck Graduate Center with the University of Mainz (MPGC), and the Austrian Science Fund (FWF) via Grants P20549-N19, P22125-B12, and W1221. T.C.S. was in part supported 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

ACKNOWLEDGMENTS

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).

REFERENCES

1. Dunker, A. K.; Obradovic, Z., The protein trinity--linking function and disorder. *Nat Biotechnol* **2001**, 19 (9), 805-6.
2. Tompa, P., Intrinsically unstructured proteins. *Trends Biochem Sci* **2002**, 27 (10), 527-33.
3. Uversky, V. N., Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* **2011**, 43 (8), 1090-103.
4. Bellahcene, A.; Castronovo, V.; Ogbureke, K. U.; Fisher, L. W.; Fedarko, N. S., Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer. *Nat Rev Cancer* **2008**, 8 (3), 212-26.
5. Konrat, R., The meandering of disordered proteins in conformational space. *Structure* **2010**, 18 (4), 416-9.
6. Uversky, V. N.; Dunker, A. K., Understanding protein non-folding. *Biochim Biophys Acta* **2010**, 1804 (6), 1231-64.
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 (36), 15138-48.
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 (2), 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 the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **2010**, 18 (4), 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 (2), 359-74.
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 (6), 2351-66.
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 (5), 1015-26.
13. Neduva, V.; Russell, R. B., Linear motifs: evolutionary interaction switches. *FEBS Lett* **2005**, 579 (15), 3342-5.
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. *Protein Sci* **2013**, 22 (3), 258-73.
15. Jensen, M. R.; Ruigrok, R. W.; Blackledge, M., Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* **2013**, 23 (3), 426-35.
16. Platzer, G.; Schedlbauer, A.; Chemelli, A.; Ozdowy, P.; Coudevylle, N.; Auer, R.; Kontaxis, G.; Hartl, M.; Miles, A. J.; Wallace, B. A.; Glatter, O.; Bister, K.; Konrat, R., The metastasis-associated extracellular matrix protein osteopontin forms transient structure in ligand interaction sites. *Biochemistry* **2011**, 50 (27), 6113-24.

17. Holde, K. E. v.; Johnson, W. C.; Ho, P. S., Principles of Physical Biochemistry. *Prentice Hall, Upper Saddle River, NJ* **1998**.
18. Uversky, V. N., Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **2002**, *11* (4), 739-56.
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* (25), 8657-68.
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* **2012**, *8* (1), 58-68.
21. Drescher, M., EPR in Protein Science Intrinsically Disordered Proteins. *Epr Spectroscopy: Applications in Chemistry and Biology* **2012**, *321*, 91-119.
22. Jeschke, G.; Polyhach, Y., Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys Chem Chem Phys* **2007**, *9* (16), 1895-910.
23. Clore, G. M.; Tang, C.; Iwahara, J., Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr Opin Struct Biol* **2007**, *17* (5), 603-16.
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* (41), 20596-608.
25. Junk, M. J.; Spiess, H. W.; Hinderberger, D., The distribution of fatty acids reveals the functional structure of human serum albumin. *Angew Chem Int Ed Engl* **2010**, *49* (46), 8755-9.
26. 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.; Kay, L. E., Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **1994**, *33* (19), 5984-6003.
27. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995**, *6* (3), 277-93.
28. Means, G. E.; Feeney, R. E., Reductive alkylation of amino groups in proteins. *Biochemistry* **1968**, *7* (6), 2192-201.
29. Schweiger, A.; Jeschke, G., Principles of pulse electron paramagnetic resonance. *Oxford University Press, New York* **2001**.
30. Tanford, C., Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions. *Journal of the American Chemical Society* **1964**, *86* (10), 2050-2059.
31. Sinha, N.; Smith-Gill, S. J., Electrostatics in protein binding and function. *Current Protein & Peptide Science* **2002**, *3* (6), 601-614.
32. Avbelj, F.; Fele, L., Role of main-chain electrostatics, hydrophobic effect and side-chain conformational entropy in determining the secondary structure of proteins. *Journal of Molecular Biology* **1998**, *279* (3), 665-684.
33. Candotti, M.; Esteban-Martin, S.; Salvatella, X.; Orozco, M., Toward an atomistic description of the urea-denatured state of proteins. *Proc Natl Acad Sci U S A* **2013**, *110* (15), 5933-8.
34. Yang, Z.; Xiu, P.; Shi, B.; Hua, L.; Zhou, R., Coherent microscopic picture for urea-induced denaturation of proteins. *J Phys Chem B* **2012**, *116* (30), 8856-62.

35. Rubinstein, M.; Colby, R. H., Polymer Physics. *Oxford University Press, New York* **2003**.
36. 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. *Journal of the American Chemical Society* **2012**, *134* (9), 4429-4436.
37. 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* (5), 1341-51.
38. Mittag, T.; Forman-Kay, J. D., Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **2007**, *17* (1), 3-14.
39. Uversky, V. N.; Santambrogio, C.; Brocca, S.; Grandori, R., Length-dependent compaction of intrinsically disordered proteins. *FEBS Lett* **2012**, *586* (1), 70-3.
40. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **1999**, *293* (2), 321-31.
41. Tompa, P.; Fuxreiter, M., Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **2008**, *33* (1), 2-8.
42. Koshland, D. E., Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **1958**, *44* (2), 98-104.
43. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **2009**, *5* (11), 789-96.
44. Grunberg, R.; Leckner, J.; Nilges, M., Complementarity of structure ensembles in protein-protein binding. *Structure* **2004**, *12* (12), 2125-36.
45. 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 U S A* **2012**, *109* (40), 16155-60.

SUPPORTING INFORMATION TO CHAPTER II

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.

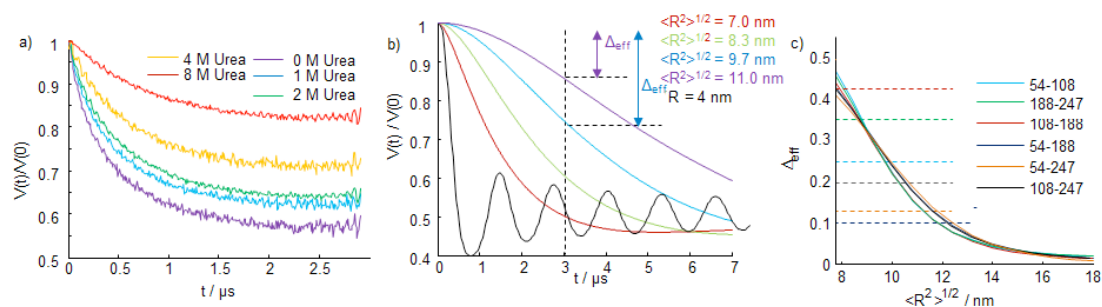


Figure S1. a) DEER time traces of C108-C188 at different urea concentrations. b) Calculated DEER time traces for different interspin distributions distances based on random coils with different $\langle R^2 \rangle^{1/2}$. Δ_{eff} is defined as the signal decay at $t = 3 \mu\text{s}$, as indicated by 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_{∞} .

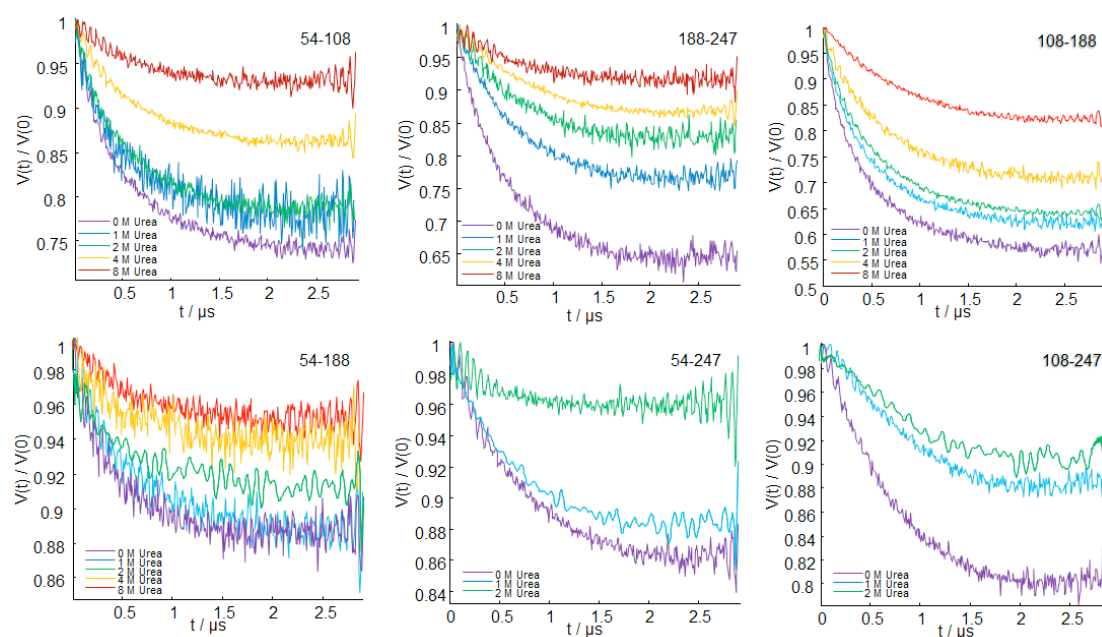


Figure S2. DEER time domain data for all double mutants for different urea concentrations. Absence of time traces indicates that we observed $\Delta_{\text{eff}} = 0$. Note the different scales of the $V(t)/V(0)$ axes.

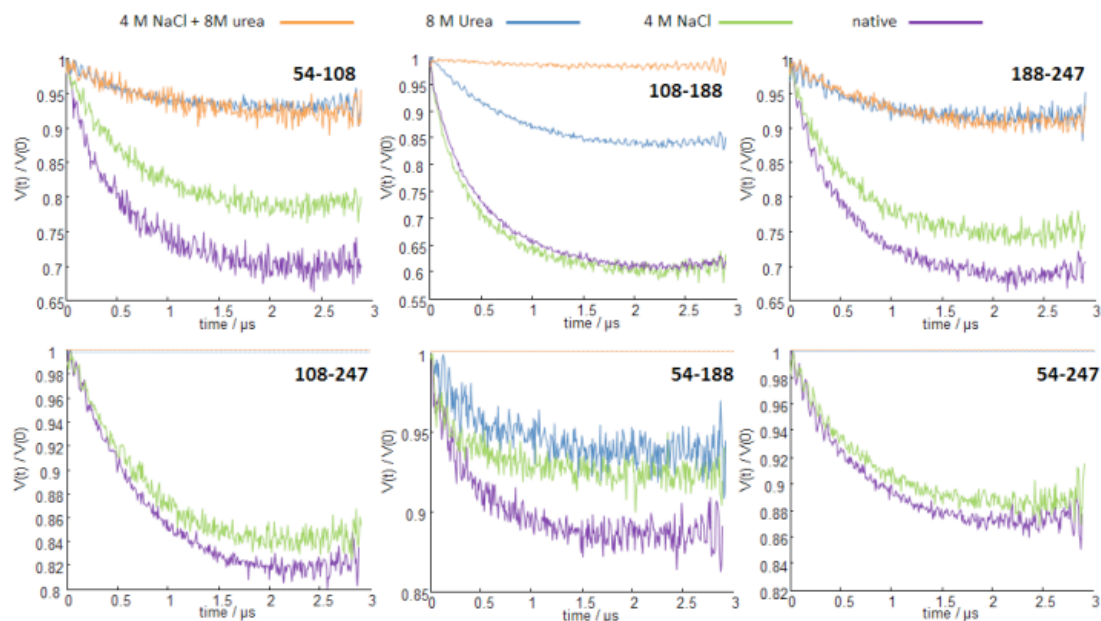


Figure S3. DEER time domain data for all double mutants for different denaturation conditions. Absence of time traces indicates that we observed $\Delta_{\text{eff}} = 0$. Note the different scales of the $V(t)/V(0)$ axes.

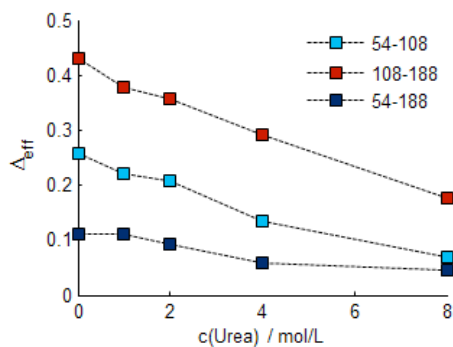


Figure S4. Δ_{eff} for several double mutants as a function of urea concentration.

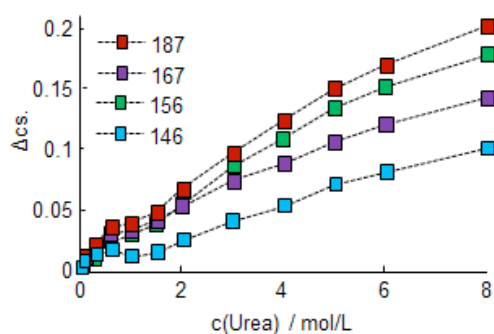


Figure S5. NMR chemical shift changes of selected residues as a function of urea concentration. Note that residue 146 shows a rather different behavior than the other three residues. However, like residue 146 also residues 156 and 167 are located in the central, more compact region of OPN.² Only 187 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:³

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

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 $V(3\mu s) = 1 - \Delta_{\text{eff}}$ 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 hypothetical fraction of compact conformations of the overall conformational ensemble. Therefore, $R(x)$ has to fulfill:

$$x = \int_{R_{\min}}^{R(x)} P(R, n, l) dr / \int_{R_{\min}}^{R_{\max}} P(R, n, l) dr \quad (\text{see Figure S6a})$$

$R_{\min}$ is depending on the pump pulse length, τ_P , of a DEER experiment. For $\tau_P = 12$ ns $R_{\min} = 1.6$ nm.⁴ $R_{\max}$ is approximately 40 nm as calculated by Jeschke and co-workers.⁵ For $x = 0.05$, that is the cooperatively folded states are estimated to the most compact 5 % of $P(R,n,l)$ and for $t = 3$ μ s eq. S6 yields the function plotted in Figure S6b.

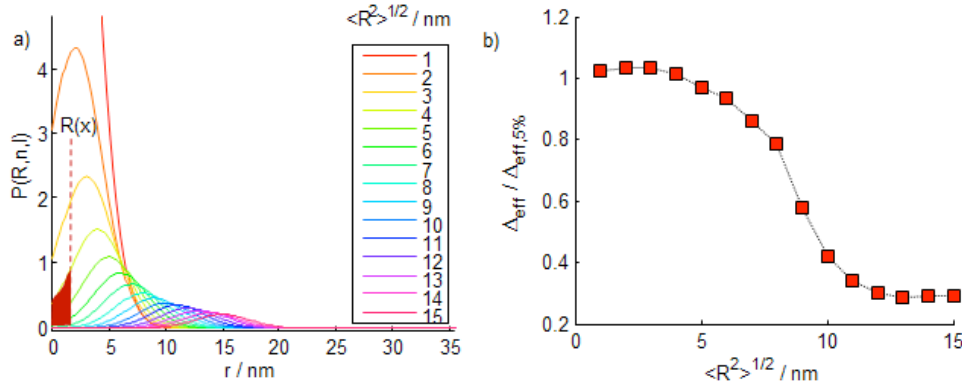


Figure S6. a) End-to-end vector distributions of random coils for different $\langle R^2 \rangle$. $R(x)$ is defined as the distance at which the integral between $R_{\min}$ and $R(x)$ matches $x\%$ of the total integral of $P(R,n,l)$. c_∞ was assumed as 2. b) Inverse contribution of the most compact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

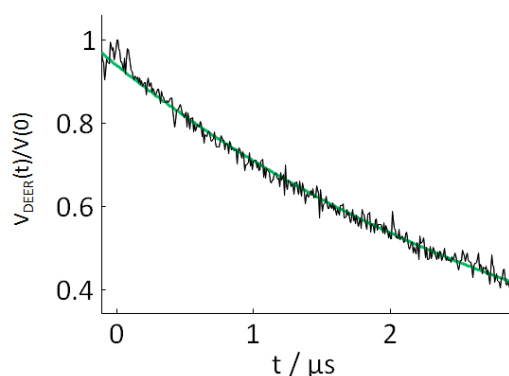


Figure S7. Normalized, raw DEER time domain data for a 0.8 mM solution of single mutant C188 (representative also for C54, C108 and C247; these single mutant data were used for experimental background correction). The green fit corresponds to a homogeneous ($d = 3$) distribution of spins in the freeze-quenched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

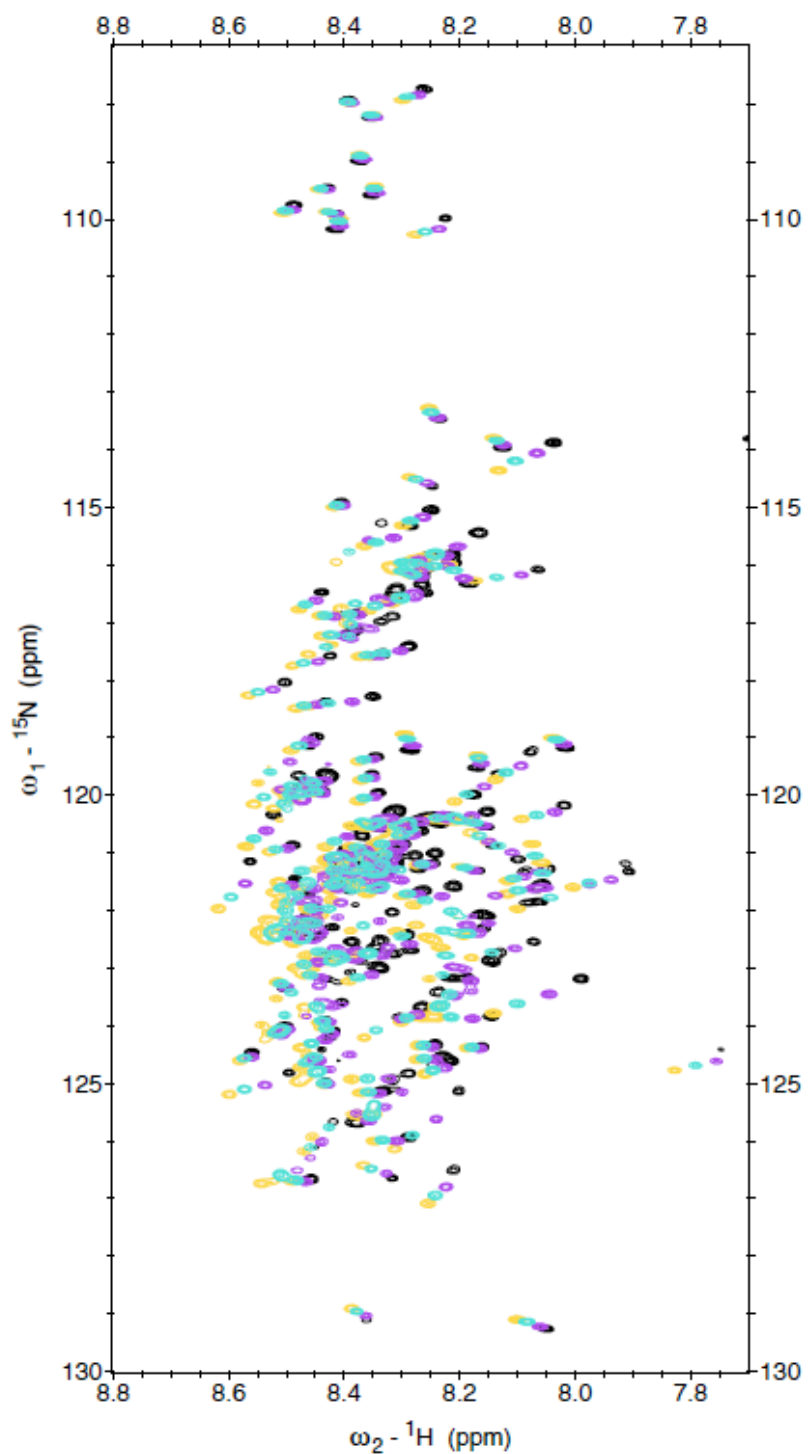


Figure S8. Full ${}^{15}\text{N}$ - ${}^1\text{H}$ HSQC spectra at different urea concentrations. Green: 0 M urea; pink: 2 M urea; blue: 4 M urea; black: 6 M urea

References:

1. Rubinstein, M.; Colby, R. H., Polymer Physics. *Oxford University Press, New York* **2004**.
2. Platzer, G.; Schedlbauer, A.; Chemelli, A.; Ozdowy, P.; Coudevylle, N.; Auer, R.; Kontaxis, G.; Hartl, M.; Miles, A. J.; Wallace, B. A.; Glatter, O.; Bister, K.; Konrat, R., The metastasis-associated extracellular matrix protein osteopontin forms transient structure in ligand interaction sites. *Biochemistry* **2011**, *50* (27), 6113-24.
3. Schweiger, A.; Jeschke, G., Principles of Pulse Electron Paramagnetic Resonance. *Oxford University Press, New York* **2001**.
4. Tsvetkov, Y. D.; Milov, A. D.; Maryasov, A. G., Pulse electron-electron double resonance (PELDOR) as nanometre range EPR spectroscopy. *Uspekhi Khimii* **2008**, *77* (6), 515-550.
5. Jeschke, G.; Koch, A.; Jonas, U.; Godt, A., Direct conversion of EPR dipolar time evolution data to distance distributions. *Journal of Magnetic Resonance* **2002**, *155* (1), 72-82.

Chapter III

Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex

Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex

Kurzbach D,* Schwarz TC,* Platzer G, Höfler S, Hinderberger D, Konrat R.

Angewandte Chemie **2013**, under revision

*both authors contributed equally

INTRODUCTION

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.^{1,2} 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.³ 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.⁴⁻⁶ The binding modes between IDPs and their ligands nonetheless remain vaguely described. Different mechanisms have been distinguished: among them so-called “folding-upon-binding”,⁷ 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.⁸ Since IDPs are typically charged higher than globular proteins, electrostatics play a pronounced role in mediating the protein-ligand interactions of an IDP.⁹ Specific charge patterns of IDPs have been found to be optimized for attraction to and interaction with their naturally charged substrates.¹⁰⁻¹² 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.¹³⁻¹⁷ We analyze the interaction of the IDP Osteopontin (OPN), an extracellular matrix protein associated with metastasis of several kinds of cancer,¹⁸⁻²¹ with heparin,¹⁷ 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 signaling and

adhesion.¹⁹ Glycosaminoglycans like heparin, hyaluronan and heparan-sulfate are crucial for chemokine function *in vivo*.^{20,21} Additionally, interactions between IDPs and biological polyelectrolytes are quite common² 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.

RESULTS

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-sulfonylthioate, 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 PREs of IDPs one should be aware that due to the rapid conformational sampling of IDPs one observes ensemble averaged PRE data.¹³ Hence, all conclusions drawn from these refer to “average” conformations and Δ PRE > 0 indicates “on average” increasing distance between labeling site and a residue upon binding, Δ PRE < 0 the opposite. The red bar in Figure 1a) indicates the heparin binding site¹⁷ (residues located in the region had to be largely excluded from the analysis because of significant signal overlap and reduced intensities in the heparin-bound state). 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, i.e., 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 mirrored in its motional dynamics. These were monitored via ^{15}N relaxation measurements (^{15}N T_1 , T_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 T_2 and ^{15}N - ^1H heteronuclear NOE differences. The regions containing most of the heparin binding site (140-170 & 180-200) display a decrease in ^{15}N T_2 due to local rigidification of the backbone upon binding (ns time scale motions). This is paralleled by larger ^{15}N - ^1H heteronuclear NOE values (Figure 1b); bottom) indicating reduced fast, picosecond timescale backbone motions. In contrast, the region encompassing residues 90-120 exhibits increased backbone flexibility as seen by both, increased T_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 validate 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.

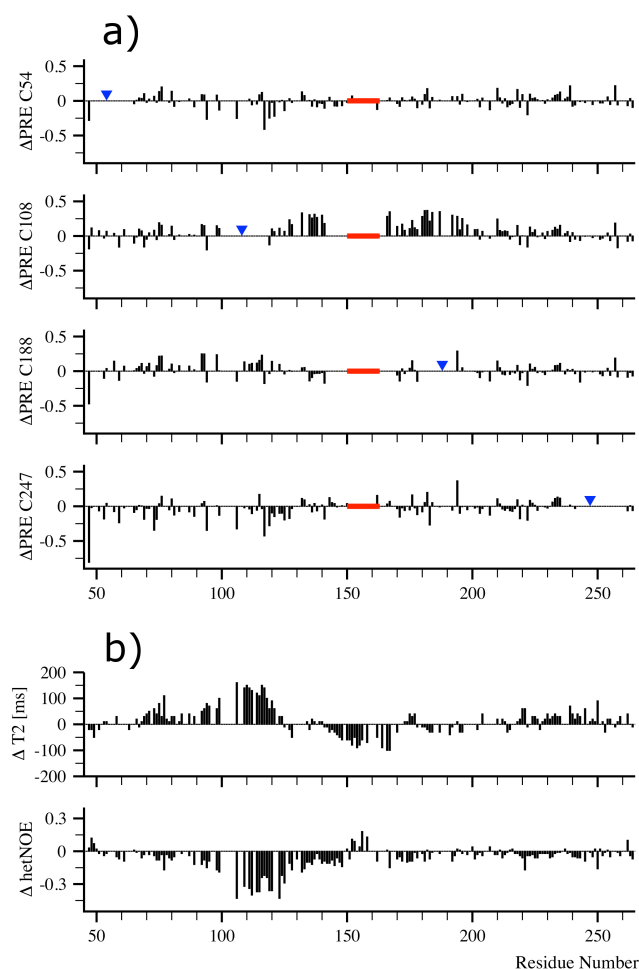


Figure 1. a) ΔPRE for the four different labeling sites upon heparin binding from ^{15}N - ^1H HSQC signal intensity changes. (ΔPRE = relative (OPN-MTSL/OPN) ^{15}N - ^1H HSQC-Intensity (heparin-bound) – relative ^{15}N - ^1H HSQC-Intensity (apo)). ΔPREs > (<) 0 indicates increasing (decreasing) distances to the labeling site (blue arrows) in the bound state. b) ΔT_2 and ΔhetNOE (^{15}N - ^1H NOE changes (bound-apo)) as a function of residue position.

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. 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 s 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.^{22,23} Note that τ_c s derived from CW EPR suffer from the non-separability 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 (T_2 and HetNOE) and EPR data are required.

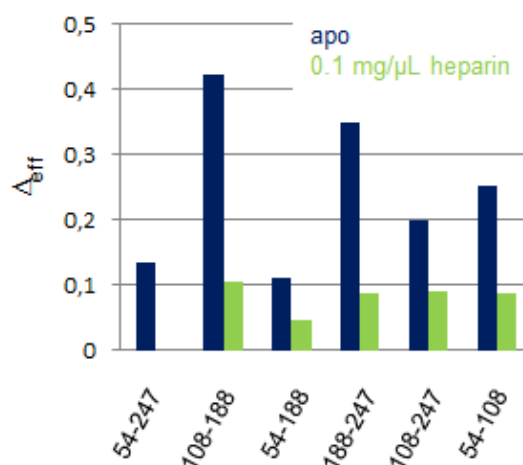


Figure 2. Δ_{eff} values for the six OPN double mutants in the apo (blue) and heparin-bound (green) state. (Δ_{eff} for mutant 54-247 with heparin is 0.)

While a detailed characterization and analysis of the DEER time traces of the OPN apo-state has already been reported,²⁴ 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.^{15,24} 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.²⁴ 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 $P(R)$ s.²⁴ 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 $P(R)$ s.²⁴ Thus, our approach is appropriate primarily for the analysis of large, extended proteins like OPN. 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.

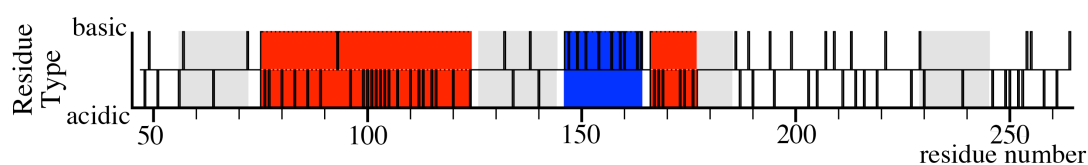


Figure 3. Charge map of OPN. Grey indicates charge neutral patches, red indicates acidic (negatively charged under experimental pH) residues and blue basic (positively charged) residues.

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 3 a charge map of OPN is shown that 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 favorable electrostatic interactions.^{17,24} Yet, the large negatively charged patch around residues 80-120 is likely to be subject to electrostatic repulsion from heparin and, therefore, to be 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.²⁴ Thus, the OPN/heparin complex reveals an electrostatic binding mode largely governed by complementary charge patterns as often found in molecular recognition processes.^{25,26} To further corroborate these findings we performed NMR chemical shift change and DPRE measurements (see Supporting Information Figure S6) and NMR diffusion oriented spectroscopy, DOSY, experiments under 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).

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 and -35 Cal/mol/K 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 with respect to the Gibbs energy of binding ($\Delta G = \Delta H - T\Delta S$). The dissociation constant was determined to 34 μ M (see the Supporting Information). Additionally, ITC measurements at different temperatures revealed a positive differential heat capacity, DC_p , (about 40 cal $M^{-1}K^{-1}$) 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).

DISCUSSION

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, rotational and translational degrees of freedom and conformational entropy, as typical for IDPs.^{9,27}

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.^{28,29} In this study conformational averaging of Sic1 in the bound state was shown to be relevant for polyvalent interactions that lead to ultra-sensitivity and non-linear binding in response to, e.g., phosphorylation. Although the details of binding differ (OPN/heparin: bulk electrostatic patches vs. 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),³⁰ 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.

ACKNOWLEDGMENTS:

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

REFERENCES:

1. Tompa, P., Intrinsically unstructured proteins. *Trends Biochem Sci* **2002**, 27 (10), 527-33.
2. Dyson, H. J.; Wright, P. E., Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **2005**, 6 (3), 197-208.
3. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **2009**, 5 (11), 789-96.
4. Markwick, P. R.; Malliavin, T.; Nilges, M., Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput Biol* **2008**, 4 (9), e1000168.
5. Montelione, G. T.; Zheng, D.; Huang, Y. J.; Gunsalus, K. C.; Szyperski, T., Protein NMR spectroscopy in structural genomics. *Nat Struct Biol* **2000**, 7 Suppl, 982-5.
6. Ozenne, V.; Schneider, R.; Yao, M. X.; 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. *Journal of the American Chemical Society* **2012**, 134 (36), 15138-15148.
7. Dyson, H. J.; Wright, P. E., Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **2002**, 12 (1), 54-60.
8. Tompa, P.; Fuxreiter, M., Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **2008**, 33 (1), 2-8.
9. Uversky, V. N., Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* **2011**, 43 (8), 1090-103.
10. 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 the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **2010**, 18 (4), 494-506.
11. Konrat, R., The meandering of disordered proteins in conformational space. *Structure* **2010**, 18 (4), 416-9.
12. Iwahara, J.; Schwieters, C. D.; Clore, G. M., Ensemble approach for NMR structure refinement against H-1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *Journal of the American Chemical Society* **2004**, 126 (18), 5879-5896.
13. Clore, G. M.; Tang, C.; Iwahara, J., Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr Opin Struct Biol* **2007**, 17 (5), 603-16.
14. Jeschke, G.; Polyhach, Y., Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Physical Chemistry Chemical Physics* **2007**, 9 (16), 1895-1910.
15. Drescher, M., EPR in Protein Science Intrinsically Disordered Proteins. *Epr Spectroscopy: Applications in Chemistry and Biology* **2012**, 321, 91-119.
16. Bund, T.; Boggs, J. M.; Harauz, G.; Hellmann, N.; Hinderberger, D., Copper uptake induces self-assembly of 18.5 kDa myelin basic protein (MBP). *Biophys J* **2010**, 99 (9), 3020-8.
17. Platzer, G.; Schedlbauer, A.; Chemelli, A.; Ozdowry, P.; Coudevylle, N.; Auer, R.; Kontaxis, G.; Hartl, M.; Miles, A. J.; Wallace, B. A.; Glatter, O.; Bister, K.; Konrat, R., The metastasis-associated extracellular matrix protein osteopontin forms

transient structure in ligand interaction sites. *Biochemistry* **2011**, *50* (27), 6113-24.

18. Bellahcene, A.; Castronovo, V.; Ogbureke, K. U.; Fisher, L. W.; Fedarko, N. S., Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer. *Nat Rev Cancer* **2008**, *8* (3), 212-26.

19. Ponta, H.; Sherman, L.; Herrlich, P. A., CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol* **2003**, *4* (1), 33-45.

20. Jones, M.; Tussey, L.; Athanasou, N.; Jackson, D. G., Heparan sulfate proteoglycan isoforms of the CD44 hyaluronan receptor induced in human inflammatory macrophages can function as paracrine regulators of fibroblast growth factor action. *J Biol Chem* **2000**, *275* (11), 7964-74.

21. de Paz, L. L.; Moseman, E. A.; Noti, C.; Polito, L.; von Andrian, U. H.; Seeberger, P. H., Profiling heparin-chemokine interactions using synthetic tools. *Acs Chemical Biology* **2007**, *2* (11), 735-744.

22. Hubbell, W.; Cafiso, D.; Altenbach, C., Identifying conformational changes with site-directed spin labeling. *Nature Struct. Biol.* **2000**, *7*, 735-739.

23. Hubbell, W.; Mchaourab, H.; Altenbach, C.; Lietzow, M., Watching your protein move using site-directed spin labeling. *Structure* **1996**, *4*, 779-783.

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

25. Honig, B.; Nicholls, A., Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144-1149.

26. Sinha, N.; Smith-Gill, S. J., Electrostatics in Protein Binding and Function. *Current Protein and Peptide Science* **2002**, *3*, 601-614.

27. Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J., Conformational entropy in molecular recognition by proteins. *Nature* **2007**, *448* (7151), 325-U3.

28. Borg, M.; Mittag, T.; Pawson, T.; Tyers, M.; Forman-Kay, J. D.; Chan, H. S., Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc. Nat. Acad. Sci.* **2007**, *104*, 9650-9655.

29. Mittag, T.; Orlicky, S.; Choy, W. Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D., Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci U S A* **2008**, *105* (46), 17772-7.

30. Kurzbach, D.; Hassounah, W.; McDaniel, J. R.; Jaumann, E. A.; Chilkoti, A.; Hinderberger, D., Hydration Layer Coupling and Cooperativity in Phase Behavior of Stimulus Responsive Peptide Polymers *J. Am. Chem. Soc.* **2013**, *135*, 11299-11308.

SUPPORTING INFORMATION TO CHAPTER III

Contents:

1. Supplementary Materials

1.1 Protein Preparation

1.2 DEER

1.3 NMR Spectroscopy

1.4 Paramagnetic Relaxation Enhancements (PREs)

1.5 ^{15}N $T_{1,2}$ and ^{15}N - ^1H NOE Spectra

2. Supplementary Methods

2.1 DEER

2.2 ITC

2.3 Experiments Under High Salt Conditions

1. Supplementary Materials

1.1 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% H_2O / D_2O 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.

1.2 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 iso-pentane. Such, a snapshot representative for the solution at the glass transition temperature is detected. The sample volume was always large enough to fill the complete resonator. The four pulse DEER sequence $\pi/2(\nu_{\text{obs}}) - \tau_1 - \pi(\nu_{\text{obs}}) - (\tau_1 + t) - \pi(\nu_{\text{pump}}) - (\tau_2 - t) - \pi(\nu_{\text{obs}}) - \tau_2 - \text{echo}$ 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 $\tau_2 = 2 \mu\text{s}$ and τ_1 were kept constant. Proton modulation was averaged by the addition of eight time traces of variable τ_1 , starting with $\tau_{1,0} = 200 \text{ ns}$ and incrementing by $\Delta\tau_1 = 8 \text{ ns}$. 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 $\nu_{\text{pump}} + 61.6 \text{ MHz}$, 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 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.

1.3 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 NMRPipe⁴ and analyzed using SPARKY⁵. 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.

1.4 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- Δ^3 -pyrroline-3-methyl) Methanethiosulfonate (MTSL). The overall PRE effect on OPN220 was measured as intensity ratio of cross-peaks in presence (I_{+MTSL}) and absence (I_{-MTSL}) of the cysteine-attached spin labels, as $\Delta_{MTSL} = I_{+MTSL}/I_{-MTSL}$. To account for possible interactions of the spin-label with the protein, MTSL was added to untagged, ^{15}N 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: $PRE = \Delta_{MTSL} + (1 - \Delta_{UNSPEC})$. 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, ^{14}N OPN Mutants C54 and C247 were tagged with MTSL and each mixed with ^{15}N OPN and heparin (^{14}N OPN 0.4 mM, ^{15}N OPN 0.4 mM, heparin 0.1 mg/ml) with no measurable PRE effect.

1.5 ^{15}N $T_{1,2}$ 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 SPARKY⁵ 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. 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 $^{15}\text{N}\{-^1\text{H}^{\text{N}}\}$ 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.

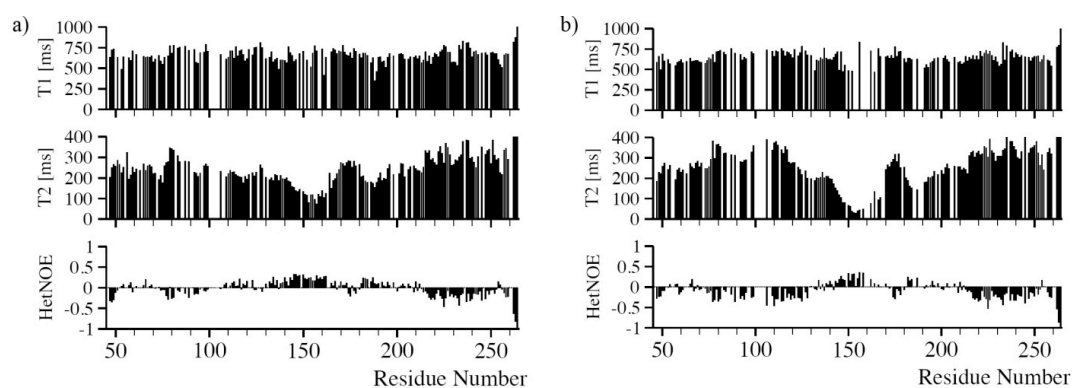


Figure S1. ^{15}N $T_{1,2}$ and $^{15}\text{N}\text{-}^1\text{H}$ NOE values for OPN in its a) free and b) heparin-bound state. Please note the differences in regions 90-120 and 140-170 in both ^{15}N T_2 times and $^{15}\text{N}\text{-}^1\text{H}$ NOE relative intensities (highlighted in Figure 1b) as ΔT_2 and ΔHetNOE).

2. Supplementary Methods

2.1 DEER

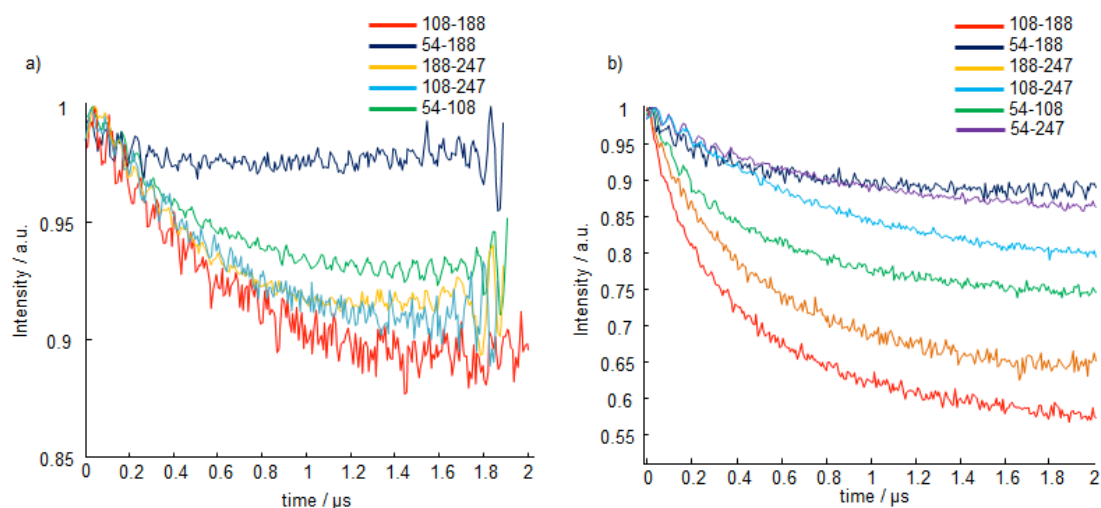


Figure S2. DEER time traces of the double mutants under investigation a) in the presence and b) in the absence of 0.1 mg/ μ L heparin. Note the different intensity scales. Missing time traces denote $\Delta_{\text{eff}} = 0$.

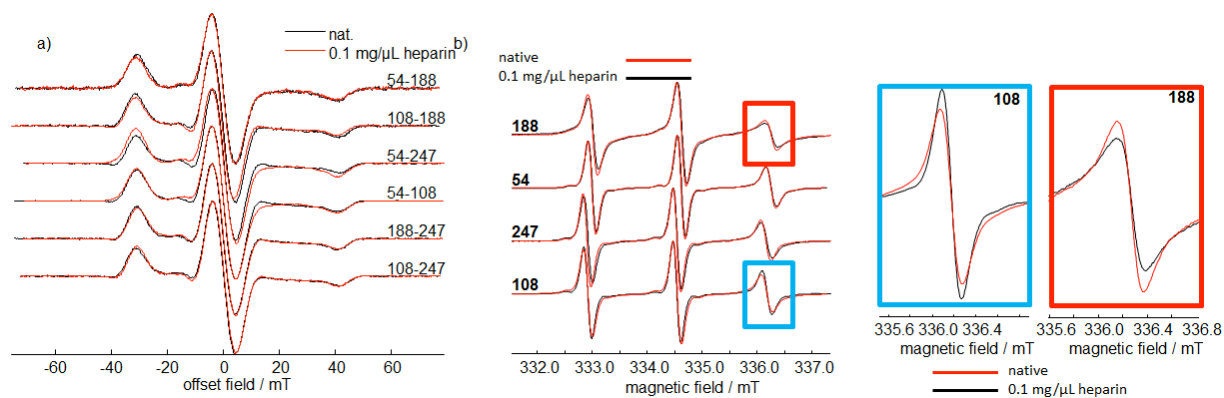


Figure S3. a) Pseudo modulated low temperature spin echoes of the double-mutants under investigation in the absence (black) and presence (red) of 0.1 mg/ μ L heparin. The modulation frequency was 3 mT. b) Room temperature CW EPR spectra of the four single-mutants under investigation in the absence (red)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

2.2 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 Microcal™ 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 μL 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 Origin™ Program⁷ 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.

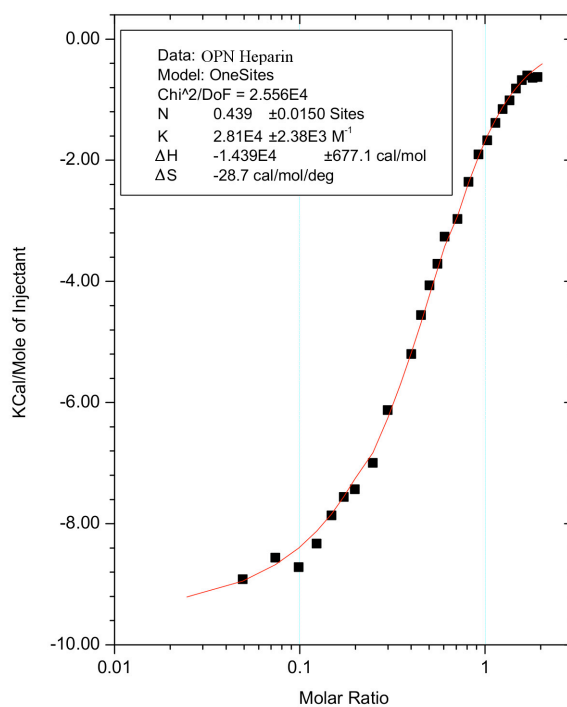


Figure S4. ITC least squares fit of incremental heat release per mole of ligand from 29 automatic injections of heparin (1.5mM) into OPN (135 μ M). The data in the inlet represents a single run (as opposed to the average of 3 given in the main text). Note that the molar ratio is plotted on a logarithmic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

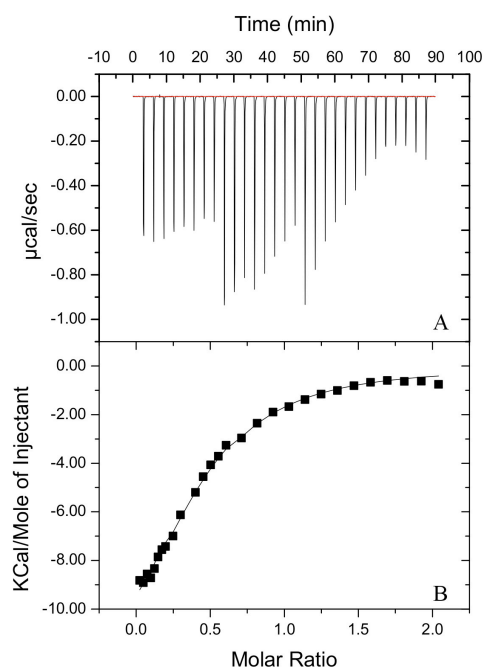


Figure S5. a) Raw data from 29 automatic injections of heparin (1.5 mM) to an OPN solution (135 μ M). b) Least square fit of incremental heat release per mole

of ligand. The obtained stoichiometry indicates 0.42 heparin strands per OPN molecule.

2.3 Experiments Under High Salt Conditions

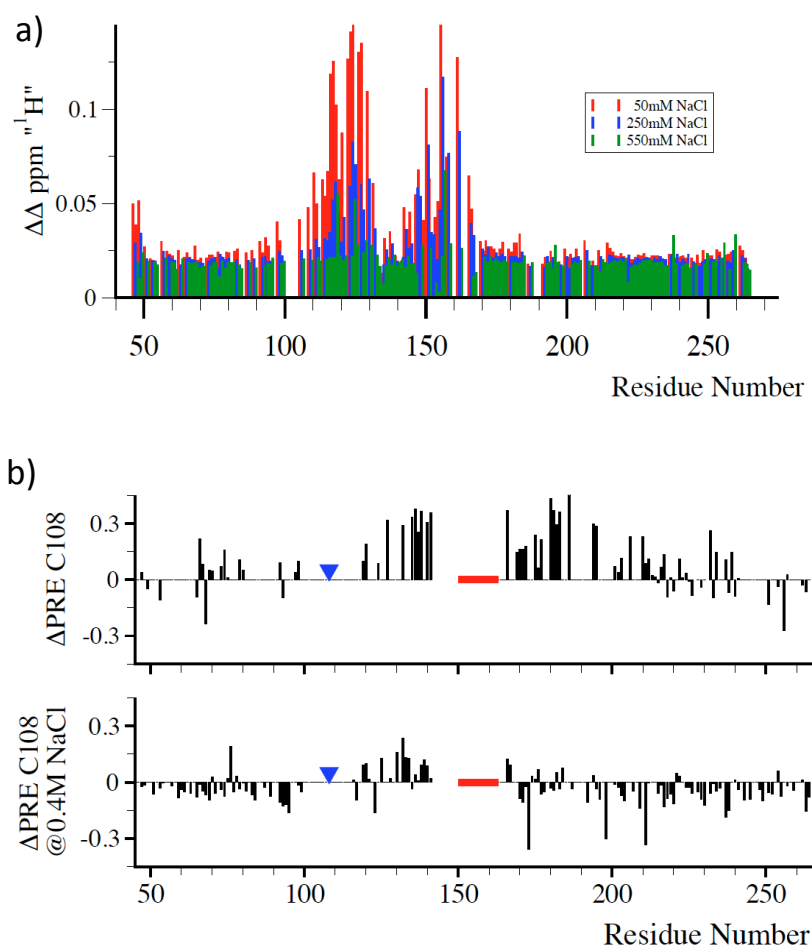


Figure S6. a) Residue plots of ΔCS : Chemical shift change of OPN ^{15}N - ^1H backbone resonances upon addition of heparin at salt concentrations of 50mM (red), 250mM (blue) and 550mM NaCl (green) (nitrogen ΔCS was divided by a factor of 5 before being combined with the proton shift to yield a “pseudo 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.

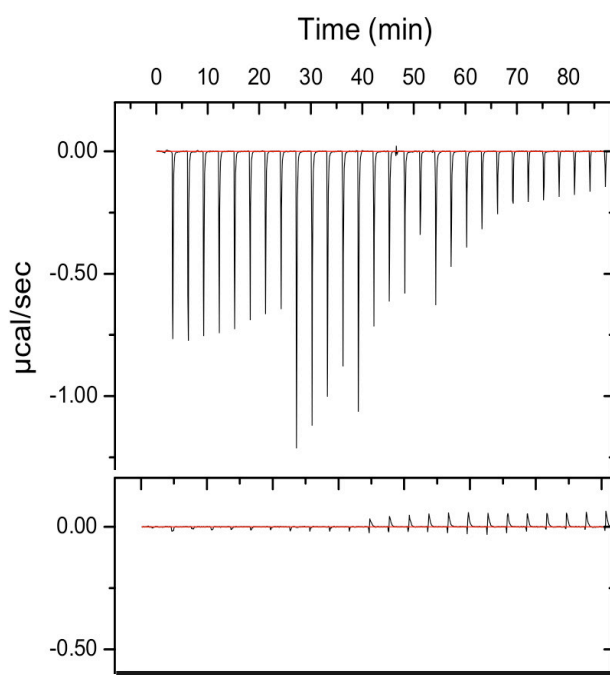


Figure S7. ITC at high salt concentrations. a) Raw-data profile for heparin (1.5 mM) titration into OPN (0.12 mM) at 50 mM NaCl. b) Raw-data obtained for a titration with the same injection scheme, temperature and similar concentrations (0.85 mM heparin, 0.1 mM OPN) at a salt concentration of 400 mM NaCl. The data was not processed nor were the measurement conditions optimized, as a different ensemble of OPN would be investigated, thus making comparison of the results 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.)

References:

1. Platzner, G.; Schedbauer, A.; Chemeli, A.; Ozdowy, P.; Coudevylle, 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.
2. 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 Edit* **2010**, *49* (46), 8755-8759.
3. 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.; Kay, L. E., Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **1994**, *33* (19), 5984-6003.
4. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **1995**, *6* (3), 277-93.
5. Goddard, T. D.; Kneller, D. G. *SPARKY 3*, University of California: San Francisco.
6. Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N., Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **1989**, *179* (1), 131-7.
7. Origin (OriginLab, N., MA).

Discussion

The biophysical characterization of IDPs is a challenging task since their dynamic nature renders conventional structure determination *via* X-Ray crystallography impossible. Two crucial questions to address in order to better understand the molecular function of an IDP are: (1) which structures are present in the conformational ensemble (and to what extent); and (2) at what timescales does the inter-conversion between these sub-structures take place. Addressing these questions however is a complicated task and several methodological challenges arise. If the number of sub-states of a protein is large the application of most solution-based methods to study IDPs faces an inherent problem since they only report on time-averaged data of the entire structural ensemble. Individual states can be detected up with CPMG-based methods under certain conditions, *i.e.*, when a species is sufficiently populated (>1%) and the inter-conversion of states takes place in the time regime between ms and μ s. Faster segmental motions, however, are in the same time-regime with the overall tumbling of the molecule and the two can no longer be separated. Consequently a direct observation of individual sub-states for IDPs is hampered in most cases. Therefore, experimental observations from different methods are usually used as constraints to construct a representative ensemble of structures that is able to reproduce the experimental observables.

In **Chapter I** we describe the application of a multitude of biophysical techniques to gain insight into the structural and dynamic properties of the protein OPN. Initial NMR observations by Fisher *et al.* already suggested, that OPN lacks a stable structure in solution based on the narrow dispersion of ^1H -chemical shifts.¹ This observation is readily verified by our data using both SAXS and CD spectroscopy. The SRCD spectra of OPN shows all characteristics of an extended polypeptide in solution and deconvolution of the data estimates the percentage of secondary structure elements to be low (5% α -helices and 13% β -sheet). Additionally, the pair correlation function obtained from SAXS measurements reveals a profile that deviates highly from a stably folded protein with maximum scattering distances ranging up to 200 Å. Both SAXS and SRCD indicate that the structural features of OPN clearly deviate from those of a well-folded protein. Closer inspection of the hydrodynamic properties of OPN led to

an interesting result. Pulsed Field Gradient (PFG) diffusion measurements determined the hydrodynamic radius of OPN to be 41 Å. By using an empirical relationship between the hydrodynamic radius (R_h) and the length of a polypeptide chain² it is possible to calculate expected values for R_h . In the case of OPN (sequence length 220 amino acids) assuming a stable fold, this would result in a hydrodynamic radius of 22.7 Å, while in the case of a complete random-coil type of structure this would yield 47.8 Å. Interestingly, an experimental value of 41 Å does not support the notion of a completely unfolded polypeptide (i.e., random coil) but hints to the presence of residual compaction in the structural ensemble of OPN.

In order to locate areas with a preference to adopt more compact conformations NMR offers a wealth of experimental approaches. ¹³C secondary chemical shifts report on deviations from random coil values on a per-residue basis. Spin relaxation measurements report on dynamics on a pico- to nanosecond timescale and, hence, on the extent to which local structure influences protein dynamics. By combining information from these two approaches it is possible to locate segments along the polypeptide that have a higher preference for residual structure. In the case of OPN, a restriction of local mobility is indeed reflected in the T_2 profile. Regions in the central region of the protein show both decreased T_2 values and increased steady-state NOE ¹⁵N-¹H} attenuation factors, suggesting a more compact structural propensity for the central region of OPN. Interestingly, no milli- to microsecond dynamics can be identified with ¹⁵N CPMG dispersion experiments as evidenced by flat dispersion profiles (R_2 versus CPMG frequency) for all residues. To complement the findings from spin dynamics measurements and secondary chemical shift mapping we employed PRE-studies on OPN. With this technique it is possible to probe long-range interactions and transient contact formation. This is done by exploiting the r^{-6} distance-dependence of the dipolar coupling between the unpaired electron of a covalently attached spin label and nearby nuclei. This results in peak broadening due to the paramagnetic effect of the label. Spin labels can be conveniently attached to cysteine side-chains. IDPs generally lack cysteines in their primary sequence but they can be easily introduced via site-directed mutagenesis. For our investigation we have chosen positions 54, 108, 188 and

247 based on low compaction values derived from the sequence-based meta-structure approach.³ Judging from the PRE-profiles of mutants C108 and C188, nuclei in the central region of OPN clearly experience a more pronounced PRE-effect apparent from the decreased signal intensities found in this part of the protein. While this observation suggests a higher structure propensity for the central region of OPN the N- and C-terminal regions of the protein appear more flexible and show little contact formation to distant sites of the protein as evidenced in the PRE profiles of mutants C54 and C247. Taken together, our experimental findings clearly show that OPN is devoid of a stable tertiary structure, but houses elements of residual structure present in the central part of the protein (residues 110-180). This finding is of special interest because this segment also contains ligand interaction motifs for integrin surface receptors to which OPN is known to bind to. It should be noted that residual structure formation is a common phenomenon in the realm of IDPs and often relevant for conformational selection-like binding events. Our finding that heparin, a known ligand for OPN is also binds to this very region, additionally supports the assumption that this might also be a relevant interaction mode for OPN.

Important conclusions drawn from Chapter I:

- OPN lacks a stable fold in solution
- OPNs dynamic and structural properties clearly deviate from a pure random-coil like behavior
- the central segment (110-180) houses preformed structural elements
- the motionally restricted part co-localizes with ligand interaction sites
- the inter-conversion between OPNs substructures takes place on timescales faster than microseconds

In **Chapter II** we use a combined EPR and NMR approach to probe the structural and dynamic preferences of OPN. In Chapter I we concluded that OPNs structural features are best described as a heterogeneous ensemble of fast inter-converting structures. Both random-coil-like as well as stabilized structural elements are found to be present simultaneously and interconvert on a sub-microsecond timescale. In order to characterize the transition events that

comprise the conformational landscape of OPN we combine EPR on freeze-quenched solutions of OPN for a static snap-shot of the system in combination with liquid-state NMR that reports on the dynamic properties of the protein.

The conformational stability of individual segments of OPN was probed by recording double electron-electron resonance (DEER) time traces on different double mutants of OPN. DEER experiments yield non-averaged ensemble data of the dipole-dipole interaction between the free electrons of the two spin labels. In total six double mutants (all of them different combinations of the four single mutants used in Chapter I) were under investigation: C54-C108, C54-C188, C54-C247, C108-C188, C108-C247 and C188-C247, each of them spanning different lengths of the protein. In order to probe the conformational stability of individual structural elements DEER time traces were recorded at different concentrations of urea. The analysis of DEER data is complicated by the fact that the heterogeneous conformational nature of OPN leads to no observable modulation of DEER time-traces and therefore, commonly applied analysis methods of inter-spin distances cannot be applied.⁴ Consequently, we use the effective modulation depth Δ_{eff} which denotes the signal decay at $t_{\text{max}} = 3\mu\text{s}$ (the maximal experimental evolution time) as a measure for compact structure populations present in the ensemble. The different double-mutants under investigation in our experiments show a very diverse response to increasing urea concentrations. The mutant C188-C247 gives rise to the steepest slope in our experiments and shows an exponential decrease of Δ_{eff} as a function of increasing urea concentration. Already low amounts of urea lead to an expansion of this segment and this region can thus be regarded as random-coil like with only very low structure content. Fragments C108-C247, C54-C108 and C108-C188 show a linear decrease suggesting more stable segments but still mostly extended structures being framed by the spin labels in these segments. A striking observation was made for both the mutant C54-C247 and C54-C188. Here we were able to observe a sigmoidal development of Δ_{eff} ; a clear indicator for cooperative folding and unfolding of protein conformations and unexpected for an IDP. The finding of cooperatively folded structures is very intriguing and the presence of stably folded sub-structures seems counterintuitive for two reasons. The $P(r)$ profile of OPN is very broad (as seen with SAXS) and DEER time traces

display no discernable modulation and converge to exponentials. These seemingly contradicting findings can be reconciled by ascribing both cooperatively folded and unfolded conformations to the structural ensemble of OPN, such that both contribute to the DEER signal. In this context it is important to note that the observation of these cooperatively folded conformations is only possible if the two labeling sites are sufficiently separated and, as a result, the modulation depth of the DEER-trace is dominated by folded contributions (C54-C247). For mutants with smaller separation, the contribution of extended conformations becomes more significant. (C54-C108, C108-C188, C188-C247) This results in different urea dependencies of the denaturation profiles (linear, exponential and sigmoidal) of the individual double mutants under investigation. As a consequence static structuring of individual segments can be ruled out, because the segments spanning smaller regions would have to show a sigmoidal denaturing dependence as well. This derivation is also in line with CPMG dispersion data that indicate a lack of a significantly populated species and inter-conversion between states on a timescale faster than μs . We hypothesized that stabilization of OPNs folded structure is additionally promoted by electrostatic interactions, given the enrichment in basic and acidic amino acid in its primary sequence. However, our findings from EPR show that complementary screening of both electrostatic (NaCl) and hydrophobic (urea) is necessary for the most effective expansion of OPN as seen in DEER denaturation profiles.

Subsequently, PRE measurements were employed to probe the structural stability of the central segment under both high urea and high salt conditions. The addition of NaCl had a significant effect on the interaction between residues 75-125 and regions flanking the spin-label at position 188. Closer inspection of the charge map of Osteopontin suggested that high salt concentrations might disrupt the electrostatic interactions between the negative charges present in region 75-125 and the positively charged patch located at residues 145-165. Interaction between these sites is also evidenced by the PRE profile of the single mutant C108 under native conditions where signals in the positive patch experience an extensive PRE effect from the spin label at position 108 resulting in peak intensities below the limit of detection. In contrast to NaCl, urea affects different parts of the protein and displaces regions 110-160 further away from

the spin label at position 188. In this case, it seems very likely that hydrophobic interactions between the patch in the central part of the protein (125-150) and the hydrophobic environment in the immediate surroundings of the spin label are disrupted. In summary the simultaneous interplay of both hydrophobic and electrostatic interactions promotes cooperative stabilization of compact sub-structures in the conformational ensemble of OPN.

Important conclusions drawn from Chapter II:

- both extended and cooperatively folded sub-structures are concomitantly populated by OPN
- the compact states are transiently stabilized by both hydrophobic and electrostatic interactions
- persistent static structure-elements are not part of OPNs conformational space

In **Chapter III** we conduct a detailed investigation of OPNs interaction with its cognate ligand heparin, a highly negatively charged glycosaminoglycan (GAG). Heparin, along with other GAGs like hyaluronan and heparan-sulfate are crucial for the *in vivo* function of chemokines and growth factors. Of special biological interest in this context is the interaction of OPN with CD44, a heparan sulfate proteoglycan (HSPG) that functions in cell signaling and adhesion and is a known receptor for heparin-binding growth factors and chemokines.⁵ We exploit information about structure (PRE) and dynamics (spin relaxation measurements) in combination with EPR to probe the conformational adaptations that OPN undergoes during the event of ligand binding. To do so, we make use of differential PRE (Δ PRE) measurements. Δ PRE values were calculated by subtracting peak-intensities in the heparin-bound state from the free state (in the absence of the ligand). Positive Δ PRE values indicate regions that experience less contact, while negative Δ PRE values indicate more contact with the spin label on a temporal average. Upon addition of heparin several segments of OPN experience a change in long-range contact formation. In the Δ PRE profile of C108, it can clearly be seen that residues in the region 120-200 are displaced from the spinlabel attached to position 108. The same observation

was made for regions 70-120 in the case of the C188 profile. Evidently, the protein experiences an expansion upon binding to heparin. This structural rearrangement is also apparent in the changes of motional dynamics. The ΔT_2 plot shows increased rigidity of the protein backbone spanning residues 140 to 170, (decreased T_2 values) and a higher flexibility for residues 70-120. This observation is mirrored in the ΔhetNOE plot. The region 150-160 is found to experience an increase in ^{15}N - ^1H heteronuclear NOE values (decreased local mobility) whereas region 70-150 becomes more flexible. Overall, these data show that Osteopontin undergoes distinct dynamic adaptations upon heparin binding.

In order to complement our findings from NMR we employed both room temperature CW-EPR on the four single mutants and DEER measurements on the six unique double mutants of OPN. Using CW-EPR we show that the spin label at position 108 gains mobility upon heparin binding whereas mutant C188 experiences a restriction in motion of the attached MTSL-label. Spinlabel motions at position 54 and 247 are not affected to a significant extent. DEER measurements were already conducted on OPN in the absence of heparin. (see Chapter II) The effect of heparin-binding on the Δ_{eff} values of the six double mutants leads to a pronounced decrease of the effective modulation depth in the DEER time traces of all six constructs under investigation. This evidence further supports our model of expansion of OPN upon ligand binding. The effect of heparin-binding can be additionally rationalized by incorporating data from chemical shift mapping. (see chapter I) Here, binding of heparin mainly affects the chemical environment of two distinct regions of the protein, 110-140 and 140-170. In conjunction with differential dynamics, ΔPRE and EPR measurements from Chapter III these data suggests that heparin directly binds to the area surrounding the positively charged patch (140-170). This is accompanied by a decrease in local mobility in this area as seen in the ΔT_2 and the ΔhetNOE plots. The second area that is highly affected (110-140) as evidenced by chemical shift mapping contains a high density of negative charges. Charge repulsion between this negative patch and the highly negatively charged heparin seems to result in a displacement of this area from the binding site,

ultimately leading to an expansion of the polypeptide. This is also evident in the dynamics data and highlighted by the increased mobility for the region 110-140.

Finally, heparin binding to OPN was monitored with isothermal titration calorimetry (ITC). Assuming an average molecular weight of 17.5kDa for OPN we obtain ΔH of -16.3 kCal/mol and $T\Delta S$ of -10.3 kCal/mol at 293K. The entropic penalty that accompanies OPN binding to its ligand is compensated by a favorable enthalpic term. ΔH and $T\Delta S$ nearly cancel out with respect to the free energy of binding resulting in a moderately strong dissociation constant of 34 μM .

Conclusion drawn from Chapter III:

- heparin binding leads to an expansion of OPN in the bound state
- OPN remains highly flexible when bound to heparin
- this conformational dynamics compensate the entropic penalty of binding

Concluding Remarks

Our detailed biophysical investigation of the IDP Osteopontin demonstrates that still a lot of misconceptions exist concerning the description of highly flexible proteins. Although a lack of persistent tertiary structure is easily determined, the dynamic and structural subtleties that accompany the dynamic nature of an IDP are very difficult to assess and cannot be described by classical models such as random-coil or molten globule. As this thesis demonstrates a combination of well-established biophysical techniques, most importantly NMR and EPR, is uniquely suited to tackle this scientific challenge. From a structural point of view, IDPs take advantage of an intricate balance between stabilizing enthalpic and destabilizing entropic contributions endowing them with a substantial conformational flexibility. In the case of OPN this flexibility results in an extremely diverse structural ensemble. We find that OPN simultaneously adopts extended as well as cooperatively folded states in solution. The latter constitute preformed structural elements that are stabilized by both hydrophobic and electrostatic long-range interactions and overlap with ligand interaction sites. Thus, exposed binding sites and preformed interaction surfaces can transiently form and consequently be recognized by cognate targets in conformational selection type of events. Our study of the interaction between OPN and its ligand heparin finds that the protein experiences significant structural and dynamical adaptations upon binding. Since the motional freedom of IDPs in solution is large, binding to these proteins is coupled with an entropic penalty due to a rigidification in the bound state. In the case of OPN this rigidification is partly compensated by a gain in flexibility of regions distant from the binding site. In conjunction with a favorable enthalpic contribution this allows for the overall free energy of binding to be favorable. We anticipate that this mode of interaction is key for disordered proteins that participate in signaling networks. In this context, a high structural plasticity is crucial in order to accommodate interactions with many different binding partners and to modulate the lifetime of the bound state by fine-tuning the entropic contribution to the free energy of binding.

REFERENCES

1. Fisher, L. W.; Torchia, D. A.; Fohr, B.; Young, M. F.; Fedarko, N. S., Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochem Biophys Res Commun* **2001**, 280 (2), 460-5.
2. Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J., Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999**, 38 (50), 16424-31.
3. Konrat, R., The protein meta-structure: a novel concept for chemical and molecular biology. *Cell Mol Life Sci* **2009**, 66 (22), 3625-39.
4. Jeschke, G.; Panek, G.; Godt, A.; Bender, A.; Paulsen, H., Data analysis procedures for pulse ELDOR measurements of broad distance distributions. *Applied Magnetic Resonance* **2004**, 26 (1-2), 223-244.
5. Jones, M.; Tussey, L.; Athanasou, N.; Jackson, D. G., Heparan sulfate proteoglycan isoforms of the CD44 hyaluronan receptor induced in human inflammatory macrophages can function as paracrine regulators of fibroblast growth factor action. *J Biol Chem* **2000**, 275 (11), 7964-74.

Acknowledgments

I want to dedicate the last few pages of my dissertation to people who supported me along the exciting journey of my PhD.

First of all I would like to thank my supervisor **Robert Konrat** for the opportunity to join his lab in the first place. I'm especially grateful for the opportunities I was given to spend time abroad. The support you gave me in both scientific and life related issues is of invaluable importance to me.

I appreciated the input and feedback I received from my PhD committee members **Gang Dong**, **Martin Tollinger** and **Bojan Zagrovic**.

I am highly indebted to **Ulrike** for her endless administrative efforts.

Big ups for all the **members** of the awesome lab I was allowed to be part of. Thanks to **Karin** for maintaining such a wonderful lab environment, and to **Georg**, **Nico** and **Dennis** for their valuable scientific input. This place became a second home for me and I enjoyed every minute I spent with you guys. A lot of memories will remain attached to this place.

I had a wonderful time during my stay in Vancouver. Thanks to **Lawrence** and **Mark** for teaching me so much about NMR and **Cecilia** and **Adrienne** for making my stay an unforgettable experience and for proofreading my thesis.

A big "thank you" to all of my friends who shared their meat and mead with me: Captain **Raphael** for his unconditional friendship over all those years; **Schicki**, probably the most altruistic person I met in my whole life, for his helpful assistance ever since the beginning of HTL; **Astrid** for all the fun times throughout my studies. Thanks for all the borrowed manuscripts. If not for you my masters would have taken years longer; **Leo**, for all the awesome time we spent together and for being generally awesome; **Georg** and **Oli** for their hospitality and for having an open ear when times were rough. Thank you so much guys; **Lisa** for your friendship over so many years; **Andreas** and **Phil** for years of excitement and adventure; **Roli** for countless pwnings and lots of recreational fun; **HC**, **Dani**, **Christoph**, and **Benj** for making my youth an unforgettable experience.

Magda, your loving and explorative nature is a limitless source of energy for me. I cherish every minute by your side.

In the end I would like to use this opportunity to thank my family and relatives. **Heli**, I am indefinitely grateful for your support over all these years. I'm glad to

have you as my uncle and friend; **Oma, Opa** and **Urli**. I can't put to words how important you are to me; **Eva**, and **Eduard** especially for their culinary assistance in both liquid and solid form; My wonderful sister **Regina**, I couldn't wish for a better sibling than you. And finally, I would like to express my deepest love for my parents **Erich** and **Roswitha**. You were always supporting my ambitions and without you I wouldn't be writing these lines. Thank you.

Curriculum Vitae

Personal Information

Nationality: Austria
Date of Birth: 27.09.1982
Place of Birth: Graz

Education

2003-2009 University of Vienna, Vienna, Austria
(Master of Science - Molecular Biology)
2010-present University of Vienna, Vienna Austria
(PhD - Molecular Biology)

Work Experience

02,03/2007 MFPL – Group Hirt, Vienna, Austria
Cell Signaling
08,09/2007 University of Vienna – Group Stuessy, Vienna, Austria
Evolutionary Genetics
08,09/2008 MFPL – Group Schlöglhofer, Vienna, Austria
Chromosome Biology
2008 - 2009 MFPL – Group Konrat, Vienna, Austria
Diploma Thesis
09-11/2012 UBC - Group McIntosh, Vancouver, Canada
NMR
03-06/2013 FMP – Group Selenko, Berlin, Germany
NMR

Languages
German: native
English: fluent
Spanish: basic

Conferences and Meetings

02/2009	Poster Presentation – MMCE2009, Otocec, Slovenia
08/2010	Poster Presentation – XXIVth ICMRBS, Cairns, Australia
09/2011	Oral Presentation - 14. Österreichische Chemietage, Linz, Austria
02/2013	Oral Presentation - MMCE2013, Semmering, Austria

Publications

The Metastasis-Associated Extracellular Matrix Protein Osteopontin Forms Transient Structure in Ligand Interaction Sites

Platzer G, Schedlbauer A, Chemelli A, Ozdowy P, Coudevylle N, Auer R, Kontaxis G, Hartl M, Miles AJ, Wallace BA, Glatter O, Bister K, Konrat R.

Biochemistry **2012** 50(27), 6113-6124

Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin

Platzer G, Kurzbach D, Schwarz TC, Henen MA, Konrat R, Hinderberger D.

Biochemistry **2013** 52(31), 5167-5175

Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex

Kurzbach D, Schwarz TC, Platzer G, Höfler S, Hinderberger D, Konrat R.

Angewandte Chemie **2013**, under revision