universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Unravelling the complexity of order within protein disorder: Osteopontin compaction, binding and phosphorylation

verfasst von / submitted by

Clara Conrad-Billroth, BSc

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

Master of Science (MSc)

Wien, 2018 / Vienna 2018

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

Acknowledgements

Firstly, I would like to thank my supervisor, Robert Konrat, who first sparked my interest in structural biology. His continued support not only made this thesis possible, but also enabled my internship at the University of Barcelona. His enthusiasm and optimism create an inspiring and positive working environment.

I am especially grateful to Borja Mateos for guiding me through this Master Thesis, always taking the time to answer questions and critically discussing my work from different angles. He has taught me a great deal while making me feel more like a colleague than his student.

Sincere thanks go to Marco Sealey, who not only helped me to develop completely new skills, but also encouraged me to be more confident in my abilities and motivated me to pursue a PhD. His curiosity, motivation and positive personality make him an amazing teacher and colleague.

Last but not least, I would like to thank Karin Ledolter and Thomas Schwarz for thought-provoking conversations during coffee breaks and their consistent support in the lab. Their patience and advice are valuable for any student working with the group.

Abstract

Osteopontin (OPN) is an intrinsically disordered, multifunctional protein recognised as a key player in a plethora of physiological and pathological processes. The extracellular matrix protein is subject to various post-translational modifications and interacts with different cell-surface receptors. There is strong evidence that OPN dynamically samples compact states, which are likely relevant for its versatile biological functions as the intricate interplay between extended and compacted conformations facilitates interactions with variable binding partners.

In this thesis, paramagnetic relaxation enhancement (PRE) was utilised to characterise the compact states of both quail and human OPN by introducing spin labels at various positions of the proteins and analysing the resulting PRE profiles. The sequence determinants for compaction in OPN were studied and the conservation of biophysical properties from the primary sequence determined. Furthermore, phosphorylation of OPN and interaction studies with its binding partners CD44, hyaluronic acid and heparin were conducted. This work paves the way for further characterisation of OPN binding to membrane-receptors and extracellular components as well as understanding the importance of compaction in intrinsically disordered proteins for their biological functions.

Zusammenfassung

Osteopontin (OPN) ist ein intrinsisch ungeordnetes Protein mit zentralen Funktionen in diversen physiologischen und pathologischen Prozessen. Es liegt als stark posttranslational modifiziertes Protein vor und ist ein Bestandteil der extrazellulären Matrix, wo es mit verschiedenen Zelloberflächenrezeptoren und anderen Biomolekülen interagiert. Es wurde gezeigt, dass OPN in der Lage ist, kompakte Strukturen auszubilden. Diese sind vermutlich essentiell für die unterschiedlichen biologischen Funktionen des Proteins, da das dynamische Zusammenspiel zwischen verschiedenen Konformationen die Interaktion mit vielfältigen Bindungspartnern ermöglicht.

In dieser Masterarbeit wurde die Kernspinresonanzmethode "Paramagnetic Relaxation Enhancement" (PRE) eingesetzt, um die kompakten Strukturen von OPN zu charakterisieren. Hierfür wurden Spin Labels an verschiedenen Positionen angebracht und die resultierenden PRE Profile analysiert. Es wurden weiterhin Sequenzdeterminanten für Kompaktierung bestimmt und die Konservierung biophysikalischer Eigenschaften analysiert. OPN wurde in-vitro phosphoryliert und Bindungsstudien mit CD44, Hyaluronsäure und Heparin durchgeführt. Die Resultate dieser Studien ermöglichen eine weitere Charakterisierung der Interaktionen von OPN mit Zelloberflächenrezeptoren und Komponenten der extrazellulären Matrix und tragen zu einem besseren Verständnis für die Rolle der Kompaktierung von intrinsisch ungeordneten Proteinen in deren biologischen Funktionen bei.

Contents

Introduction

Highly dynamic proteins which exist in a vast ensemble of heterogeneous structures, referred to as intrinsically disordered proteins (IDPs), have gained growing recognition in the last decade due to their biological importance and intriguing biophysical properties [1, 2]. The ability to sample a broad conformational space allows IDPs to interact with multiple binding partners and act as protein network hubs [3–5]. IDPs are especially abundant in eukaryotic organisms [6], which is likely due to their ability to facilitate the signalling mechanisms required for multicellular life [7, 8]. Tight regulation of these proteins is crucial [9, 10] and elevated protein levels of several IDPs are strongly associated with disease [11, 12].

IDPs do not adopt a random structure, but sample a distinct set of different predefined populations [13]. Most IDPs are defined by a high content of structure-breaking, polar and charged amino acids, while containing a low amount of bulky hydrophobic amino acids [14]. Although this means IDPs typically lack a hydrophobic core, varying degrees of compaction are possible [15], which is primarily modulated by the overall charge and proline-content of the protein [16].

Adding to the flexible nature of IDPs, several post-translational modifications (PTMs) and splicing variants increase the complexity of their conformational heterogeneity in-vivo. As they do not adopt a rigid structure, most amino acids on the sequence are usually accessible for PTMs and disordered regions tend to be targets for covalent modifications like phosphorylation [17, 18]. PTMs can modulate the function of IDPs on several levels - they are able to affect the propensity to form secondary structural elements, regulate intra- and intermolecular tertiary contacts and have a strong impact on the overall physico-chemical properties of the protein (e.g. electrostatic and hydrophobic changes), affecting the conformational ensemble sampled [19]. Disordered regions and proteins containing protein-binding motifs and PTM sites are frequently subject to tissue-specific splicing [20, 21], which could provide a mechanism for tissue-specific regulation of interaction networks.

The protein this thesis is focussed on, Osteopontin (OPN), is a multifunctional protein, expressed by various tissues and cell types [22] and part of the small integrin-binding ligand N-linked glycoprotein (SIBLING) family - soluble, secreted, integrin-binding proteins that have important functions in bone mineralisation [23] and implications in tumour-progression [24]. Epithelial cells secrete OPN into bodily fluids, including milk [25], blood [26] and urine [27].

OPN is a protein with versatile physiological functions [28] and has been shown to be involved in a plethora of pathological processes and diseases, like metabolic disorders [29], immune-mediated diseases [30] and autoimmune disorders [31]. Furthermore, OPN is associated with tumour progression and growth [32], and could serve as biomarker for the diagnosis and treatment of cancer [33]. OPN has been identified as a flexible protein in solution [34] and like many IDPs, it is part of a complex interaction network with multiple binding partners. The protein is found in the extracellular matrix of mineralized tissues [35] and is involved in the regulation of hydroxyapatite crystal formation [36]. OPN also acts as a cytokine and is able to activate migration of various immune-cells [37, 38].

Through binding to cell-surface receptors, namely several integrins [39] and CD44 variants [40], OPN can mediate different biological activities like cell survival, immune regulation, cell adhesion and migration. Therefore, it is not surprising that these interactions have been associated with the metastatic potential of OPN in cancer [41]. OPN binds integrins primarily through the canonical arginine-glycine-aspartic acid (RGD) recognition sequence [39]. CD44 consists of 4 distinct regions: the hyaluronic acid binding amino-terminal domain (HABD), the stem-part with variable length due to alternative splicing, the transmembrane region and the cytoplasmic tail [42]. Although the exact binding mechanism is not clearly understood, studies have reported interactions of OPN with the flexible stem of CD44 containing the exons v3, v6 and/or v7 [43–45].

Importantly, OPN is subject to several cell-type specific modifications including several PTMs, protein cleavage, alternative translation and splicing. It is a substrate for matrix metalloproteases [46, 47] and thrombin, which exposes the cryptic integrin binding motif of OPN [48, 49] and enhances binding to several integrins [39]. Although research has primarily focussed on secreted OPN, alternative translation generates an intracellular version of OPN lacking the signal peptide [50], which also has essential biological functions [51]. The various splicing isoforms of OPN have distinct biological functions and are expressed in different contexts [52].

Especially phosphorylation and glycosylation of OPN have been investigated by several groups. For example, it has been shown that OPN phosphorylation modulates hydroxyapatite binding, which has an influence on hydroxyapatite formation and growth [53] and regulation of macrophage migration by OPN is dependent on phosphorylation [54]. Phosphorylation and glycosylation patterns have been characterised for OPN isolated from rat bone [55], murine fibroblasts [56], bovine milk [57], as well as human milk [58] and urine [59]. Recently, a Golgi casein kinase, family with sequence similarity 20 member C (Fam20c), localized within the secretory pathway was identified, which has been shown to phosphorylate proteins in the extracellular matrix of bones and teeth as well as in milk [60, 61]. The kinase mainly phosphorylates Ser-X-Glu/pSer motifs and a mutation leading to a decreased Fam20c activity has been linked to Raine syndrome, a rare osteoclerotic bone dysplasia [60, 62, 63].

Although the inherent flexibility of IDPs can make their investigation challenging, Nuclear Magnetic Resonance (NMR) based methods have proven to be especially useful and versatile by providing structural and dynamic information for a broad range of time scales and with a coverage of almost every residue in a protein [64,65]. For example, the chemical shift is highly sensitive to changes in the environment of a residue, which makes it possible to characterise ligand binding and differences in local structure of a protein [66, 67].

A NMR technique to characterise transient long-range interactions of IDPs, Paramagnetic Relaxation Enhancement (PRE), exploits the effect of an unpaired electron, which causes paramagnetic-enhanced relaxation of any nuclear spin within its proximity through dipolar interactions [68, 69]. This paramagnetic centre is introduced in the form of a tag, typically as nitroxide free radicals or lanthanides bound via a chelator, which is conjugated to a cysteine residue of the protein [70].

The magnitude of the PRE is proportional to the r^{-6} distance between the electron and nucleus, meaning the contribution from short distances will be especially large [68, 69]. Although both longitudinal and transverse relaxation rates are affected, measurement of the transverse relaxation rate is more accurate and reliable as it is more sensitive and less susceptible to cross-correlation effects [68]. The PRE rate (Γ_2) is obtained by calculating the difference between the transverse relaxation rates of the paramagnetic sample and a diamagnetic control (in our case generated by quenching the nitroxide label), which excludes relaxation mechanisms present in both states [68]. Due to the strong distance dependence and the large effect arising from the interaction between the magnetic moment of an unpaired electron and the nucleus, information on long-range contacts (up to 35 Å) and sparsely-populated states can be obtained [71–73].

Previous studies by our group have indicated the prevalence of compacted states in OPN [74,75]. This thesis is focussed on the characterisation of these compact states of quail OPN (qOPN) and human OPN (hOPN). To generate a more accurate correlation map [76], several spin labels at an interval of ca. 20-30 residues were introduced. Conserved sequence properties of OPN were analysed with the compact states in mind. To investigate the complex interaction network of hOPN, in-vitro phosphorylation with Fam20c was carried out, the phosphorylation sites mapped by NMR and mass spectrometry and binding studies conducted with the flexible stem and the HABD of CD44 as well as hyaluronic acid - a non-sulphated glycosaminoglycan consisting of D-glucuronic acid and N-acetylglucosamine disaccharide repeats and an essential part of the extracellular matrix [77] - and heparin - a negatively charged, highly sulphated glycosaminoglycan used as an anticoagulant and structurally related to heparan sulphate, a common PTM in proteoglycans [78].

Materials and Methods

Expression and Purification of Osteopontin

Quail Osteopontin (qOPN)

Cysteine mutations for MTSL labelling were introduced via site-directed mutagenesis (positions highlighted in sequence below) using a pET11d plasmid containing the respective qOPN insert as template vector. The correct insertion of the mutation was validated via sequencing of the plasmid.

Protein Sequence of qOPN (mutation sites highlighted):

MHQDHVDSQSQEHLQQ^{60}TQNDLASLQQTHYSSEENAD^{80}VPEQPDFPDVPSKSQETVDD^{100}DDDDDNDSNDTDESDEVF
TD^{120}FPTEAPVAPFNRGDNAGRGD^{140}SVAYGFRAKAHVVKASKIRK^{160}AARKLIEDDATTEDGDSQPA^{180}GLWWPKES
REQNSRELPQHQ^{200}SVENDSRPKFDSREVDGGDS^{220}KASAGVDSRESQGSVPAVDA^{240}SNQTLESAEDAEDRHSIENN^{2}
^{60}EVTR

The respective vector was transformed into the Escherichia coli (E. coli) strain BL21(DE3) phage resistant via heat-shock method. Protein expression was induced at an OD_{600} of ca. 0.8 by adding 0.4 mM IPTG. For expression of isotopically labelled protein for NMR studies, before induction the cells were harvested and pellets from 4L LB resuspended in 1L M9-Minimal Media (containing 1g/L $^{15}NH_4Cl$ and 3g/L ^{13}C Glucose for ^{15}N and ^{13}C labelling respectively). The expression was carried out at 27°C and 140rpm overnight.

The cells were harvested and resuspended in 40mL of cold PBS (2mM KH_2PO_4, 8mM Na_2HPO_4, 2.5mM KCl, 140mM NaCl, 5mM EDTA, pH 7.3), fresh DTT was added for cysteine mutants. The suspension was sonicated and cooked at 95°C for 10min. The lysate was centrifuged at 18000 rpm and an ammonium sulphate precipitation (saturation of 50%) carried out with the supernatant. The pellet was resuspended in PBS, diluted 1:2 with water to lower the salt concentration and an anion-exchange chromatography (HiTrap Q) performed. The column was equilibrated with PBS and a gradient of 30% High-Salt Buffer (PBS containing 1M NaCl) run in 20 minutes at 2 mL/min. Protein purity was confirmed via SDS-PAGE and NMR. The protein was concentrated using a centrifugal filter and the final concentration measured at 280nm.

Human Osteopontin (hOPN)

Cysteine mutations for MTSL labelling were introduced via site-directed mutagenesis (positions highlighted in sequence below) using a pETM11 plasmid containing the respective hOPN insert as template vector. The correct insertion of the mutations was validated via sequencing of the plasmid.

Protein Sequence of hOPN (after TEV-Cleavage, mutation sites highlighted):

GAMGIPVK[20]QADSGSSEEKQLYNKYPDAV[40]ATWLNPDPSQKQNLLAPQNA[60]VSSEETNDFKQETLPSKSNE[80]SHDHM
DDMDDEDDDDHVDSQ[100]DSIDSNDSDDVDDTDDSHQS[120]DESHHSDESDELVTDFPTDL[140]PATEVFTPVVPTVDTYD
GRG[160]DSVVYGLRSKSKKFRRPDIQ[180]YPDATDEDITSHMESEELNG[200]AYKAIPVAQDLNAPSDWDSR[220]GKDSYET
SQLDDQSAETHSH[240]KQSRLYKRKANDESNEHSDV[260]IDSQELSKVSREFHS[280]HEFHSHEDMLVVDPKSKEEDKHLK
F[300]RISHELDSASSEVN

The respective vector was transformed into the E. coli strain BL21(DE3) Rosetta pLysS phage resistant via heat-shock method. Protein expression was induced at an OD_{600} of ca. 0.8 by adding 0.4 mM IPTG. For expression of isotopically labelled protein for NMR studies, before induction the cells were harvested and pellets from 4L LB resuspended in 1L M9-Minimal Media (containing 1g/L $^{15}NH_4Cl$ and 3g/L ^{13}C Glucose for ^{15}N and ^{13}C labelling respectively). The expression was carried out at 28°C and 140rpm overnight.

The cells were harvested and resuspended in 40mL of cold PBS (140mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, 20mM Imidazole, pH 8), fresh DTT was added for cysteine mutants. The suspension was sonicated and cooked at 95°C for 10min. The lysate was centrifuged at 18000 rpm and the supernatant loaded onto a HisTrap affinity column. The protein was eluted with High-Imidazole Buffer (140mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, 500mM Imidazole, pH 8).

The protein was concentrated in TEV-Cleavage Buffer (140mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, 1mM DTT, 1mM EDTA, pH 8), TEV protease was added in a ratio of 1:50 and incubated overnight at 4°C under agitation. Further purification was carried out with ion-exchange chromatography (HiTrap Q). The column was equilibrated with PBS and a gradient of 50% High-Salt Buffer (PBS containing 1M NaCl) run in 30 minutes at 1 mL/min. Protein purity was confirmed via SDS-PAGE and NMR. The protein was concentrated using a centrifugal filter and the final concentration measured at 280nm.

Expression and Purification of CD44

CD44 consists of several regions: the signal-peptide (1-20), the hyaluronic acid binding domain (21-178), a linker (179-223), the variable stem region (224-649), a transmembrane domain (650-670) and a cytoplasmic tail (671-742). The stem region as well as the hyaluronic acid binding domain were purified for binding studies with hOPN.

Complete Sequence of CD44 (hyaluronic acid binding domain is coloured blue and the stem region green):

MDKFWWHAAWGLCLVPLSLAQIDLNITCRFAGVFHVEKNGRYSISRTEAADLCKAFNSTLPTMAQMEKALSIGFETCR
YGFIEGHVVIPRIHPNSICAANNTGVYILTSNTSQYDTYCFNASAPPEEDCTSVTDLPNAFDGPITITIVNRDGTRYV
QKGEYRTNPEDIYPSNPTDDDVSSGSSSERSSTSGGYIFYTFSTVHPIPDEDSPWITDSTDRIPATTLMSTSATATET
ATKRQETWDWFSWLFLPSESKNHLHTTTQMAGTSSNTISAGWEPNEENEDERDRHLSFSGSGIDDDEDFISSTISTTP
RAFDHTKQNQDWTQWNPSHSNPEVLLQTTTRMTDVDRNGTTAYEGNWNPEAHPPLIHHEHHEEEETPHSTSTIQATPS
STTEETATQKEQWFGNRWHEGYRQTPKEDSHSTTGTAAASAHTSHPMQGRTTPSPEDSSWTDFFNPISHPMGRGHQAG
RRMDMDSSHSITLQPTANPNTGLVEDLDRTGPLSMTTQQSNSQSFSTSHEGLEEDKDHPTTSTLTSSNRNDVTGGRRD
PNHSEGSTTLLEGYTSHYPHTKESRTFIPVTSAKTGSFGVTAVTVGDSNSNVNRSLSGDQDTFHPSGGSHTTHGSESD
GHSHGSQEGGANTTSGPIRTPQIPEWLIILASLLALALILAVCIAVNSRRRCGQKKKLVINSGNGAVEDRKPSGLNGE
ASKSQEMVHLVNKESSETPDQFMTADETRNLQNVDMKIGV

Stem part of CD44 (CD44stem)

A pETM11 plasmid containing the respective CD44stem insert was transformed into the E. coli strain BL21(DE3)pLysS phage resistant via heat-shock method. A pellet from 50mL of overnight culture was resuspended in 1mL LB and added to 1L M9-Minimal Media (containing 1g/L $^{15}NH_4Cl$ and 3g/L ^{13}C Glucose for ^{15}N and ^{13}C labelling respectively) and grown to an OD_{600} of ca. 0.8 at 37°C and 140rpm, before induction with 0.4 mM IPTG. Protein expression was carried out at 32°C for 4h.

The cells were harvested and resuspended in 40mL of cold Resuspension-Buffer (50mM Tris, 300mM NaCl, 6M Guanidinium chloride, pH 8). The suspension was sonicated, the lysate was centrifuged at 18000 rpm and the supernatant (diluted to 4M Guanidinium chloride) loaded onto a HisTrap affinity column. The column was washed with Washing-Buffer 1 (50mM Tris, 300mM NaCl, 4M Guanidinium chloride, pH 8), Refolding-Buffer (50mM Tris, 300mM NaCl, pH 8) and Washing-Buffer 2 (50mM Tris, 1M NaCl, 20mM imidazole, pH 8) before eluting with Elution-Buffer (50mM Tris, 300mM NaCl, 500mM imidazole, pH 8) using a gradient of 100% in 30min at 1mL/min.

To remove the His-Tag, TEV protease was added in a ratio of 1:50 and dialysed overnight in PBS (10mM NaP, 500mM NaCl, 1mM EDTA, 1mM DTT, pH 7.4) overnight at 4°C. The protein was concentrated using a centrifugal filter with 30kDa cut-off and the buffer changed using a PD-10 Desalting-Column. Protein purity was confirmed via SDS-PAGE and NMR. The protein was concentrated using a centrifugal filter and the final concentration measured at 280nm.

Hyaluronic Acid Binding Domain of CD44 (HABD)

A pETM11 plasmid containing the respective HABD insert was transformed into the E. coli strain BL21(DE3)pLysS phage resistant via heat-shock method. Protein expression was induced at an OD_{600} of ca. 0.8 by adding 0.4 mM IPTG. Protein expression was carried out at 37°C for 4h.

The cells were harvested and resuspended in 40mL of cold Resuspension-Buffer (10mM Na_2HPO_4, 1.8mM KH_2PO_4, 140mM NaCl, 2.7mM KCl, pH 7.4). The suspension was sonicated, the lysate was centrifuged and the inclusion body pellet resuspended in ice-cold Triton-Wash (50mM Tris, 100mM NaCl, 0.5% Triton X-100, 0.02% NaN_3, pH 8). The pellet was made homogeneous using a Dounce-Homogenizer. The pellet was washed several times, finally resuspended in Solubilisation-Buffer (50mM MES, 8M Urea, 0.1mM EDTA, 0.1mM fresh DTT, pH 6.5) overnight at 4°C under agitation.

The solution was centrifuged to remove insoluble material and the protein refolded in a 200-fold dilution with Refolding-Buffer (100mM Tris, 250mM L-Arginine, 2mM EDTA, 2.5mM Reduced Glutathione, 2.5mM Oxidized Glutathione, 0.5µg/mL fresh Leupeptin, pH 8.5). Refolding was carried out for 48h at 4°C before concentrating the protein using a centrifugal filter.

Monomeric protein was purified by size-exclusion chromatography (26/60 Superdex G75) in SEC-Buffer (20mM Tris, 150mM NaCl, pH 8) at 4mL/min. The protein was concentrated and TEV protease was added in a ratio of 1:50 and dialysed overnight in PBS with low imidazole (140mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, 20mM Imidazole, pH 8) at 4°C. The fractions containing the protein were loaded onto a HisTrap affinity column and the flow-through collected to remove TEV protease and the His-Tag. Protein purity was confirmed via SDS-PAGE and NMR. The protein was concentrated using a centrifugal filter and the final concentration measured at 280nm.

Phosphorylation

Fam20c Kinase

HEK293T cells, stably expressing FLAG-tagged Fam20c were kindly provided by Vincent S. Tagliabracci (UT Southwestern Medical Center, Texas). For expression and purification, the protocol was followed as described by Tagliabracci et. al [60].

The HEK cells were grown for 72h to allow expression of high amounts of Fam20c. All subsequent steps were carried out at 4°C. The media was centrifuged, filtered and incubated with anti-FLAG resin overnight. The solution was loaded onto a 20mL column and washed with Phosphorylation-Buffer (50mM Tris, 50mM NaCl, pH = 7.2). The resin was incubated with Elution-Buffer (100µg/mL FLAG Peptide in Phosphorylation-Buffer) for 4h under agitation before elution. Protein purity was confirmed via SDS-PAGE. The protein was stored at -80°C with 15% glycerol and 1mM beta-mercaptoethanol.

Phosphorylation Reaction

The reaction was carried out in Phosphorylation-Buffer (50mM Tris, 50mM NaCl, pH 7.2). The reaction mix consisted of 50µM protein, an excess of ATP (4.3mM for hOPN and 0.8mM for CD44 based on the expected phosphorylation sites), 10mM $MnCl_2$. Fam20c was added at a ratio of ca. 1:300. The reaction was carried out in 3 steps, each 8-16h, at room temperature. After each step the solution was centrifuged at 18000 rpm and fresh ATP and $MnCl_2$ added to the supernatant. The protein was purified using ion-exchange chromatography after removing the ATP through a centrifugal filter.

The phosphorylated hOPN sample was sent to the MFPL Mass Spectrometry Facility and intact protein mass determination as well as mapping of the phosphorylation sites were performed by Dorothea Anrather.

NMR Measurements

MTSL-Tagging

The protein was incubated with an excess of DTT (10mM) for 15min at room temperature. A PD-10 Desalting-Column was equilibrated with MTSL-Tagging Buffer (100mM NaP, 1mM EDTA, pH 8), the protein loaded to a total volume of 2.5mL and eluted with 3.4mL of buffer. The free thiol concentration was measured using a 300µM DTNB solution at 412nm. A 3-fold excess of MTSL was added and the sample incubated for 3 hours at 37°C during agitation. The free thiol concentration was measured again to confirm complete tagging of the protein.

Measurement Conditions

	Buffer	Temperature [K]	Concentration [mM]
qOPN	qOPN PBS, pH 6.5	298	0.5
hOPN	50mM NaCl, 50mM NaP, pH 6.5	293	0.2-0.3
hOPN (phos.)	50mM NaCl, 50mM NaP, pH 6.5	293	0.1
CD44stem	50mM NaCl, 50mM NaP, pH 6.5	293	0.1
HABD	50mM NaCl, 50mM NaP, pH 6.5	293	0.1

Table 2.1. Measurement Conditions for NMR-Samples

NMR Spectra were recorded using a 800 MHz Bruker Avance III and 600 MHz Bruker Neo spectrometer; 10% D_2O was added as the lock-solvent. For hOPN and CD44 measurements, 5mm Shigemi-Tubes were used to reduce the sample volume needed (250µL compared to 500µL for qOPN).

For the $H^N_{T\,2}$ PRE measurements, paramagnetic samples were reduced using a 3-fold excess of ascorbic acid providing the diamagnetic control-samples. Relaxation delays for hOPN were 0.001, 0.0025, 0.005, 0.015, 0.03, 0.06 and 0.12 s and for qOPN 0.001, 0.005, 0.01, 0.02, 0.05 and 0.1 s.

For interaction studies, heparin sodium salt from porcine intestinal mucosa with an average molecular weight of 17.5kDa and hyaluronic acid sodium salt from streptococcus equi with an average molecular weight of 20kDa were bought from Sigma Aldrich. The titration of hOPN and phosphorylated hOPN with hyaluronic acid were carried out in molar ratios of ca. 10:1, 5:1, 2:1, 1:1 and 1:2. Titration with heparin were carried out in molar ratios of ca. 10:1, 5:1, 2:1, 1:1, 1:2, 1:5 and 1:10.

Data Analysis

All spectra were processed with NMRPipe [79] and analysed with Sparky [80]. Further data analysis and visualisation were carried out using RStudio [81].

The Multiple Sequence Alignment of OPN was performed out using MAFFT [82] and the phylogenetic tree generated using Archaeopteryx [83].

PRE profiles and Correlation Matrix

Relaxation fitting was carried out using Sparky. The points are fit to a decaying exponential ($h = A * e^{-R*t}$) and the time constant (T) given, including an error estimate. The PRE rates were calculated as the difference between the transverse relaxation rates ($R_2 = \frac{1}{T}$) for para- and diamagnetic measurements for each residue.

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

The symmetric correlation matrix was calculated according to the pearson correlation coefficient. With the sample size (n) as the number of labelling sites and the PRE rates as sample points (x, y).

$$cov_{x,y} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

$$corr_{x,y} = \frac{cov_{x,y}}{\sigma_x \sigma_y}$$

Calculation of the Chemical Shift Perturbation (CSP)

To determine the binding site of heparin to hOPN, the CSP was calculated for each residue upon titration of heparin and plotted. The CSP is calculated as the geometrical distance between the chemical shift of a residue before and after addition of a ligand.

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\frac{\Delta\delta_N}{5})^2}$$

Phosphorylation Analysis

The phosphorylation sites from the mass spectrometry results and the NMR assignment were compared with the phosphorylation sites of hOPN found in the literature for human milk OPN [58] and OPN phosphorylated by Fam20c in HepG2 cells [61].

Plotting Hydrophobicity and Charge Distribution of Osteopontin

Hydrophobicity plots were calculated using the hydrophobicity scale of Kyte & Doolittle [84]. Charge plots were simply calculated by assuming the amino acids lysine and arginine as +1, histidine as +0.5 and glutamic acid and aspartic acid as -1. The average for every residue was calculated with a sliding window of size 15 to generate distribution plots.

Results

Sequence Properties of OPN

To investigate the sequence determinants of the compact states of OPN, the distribution of charged and hydrophobic residues was analysed. A multiple sequence alignment (MSA) was generated (supplementary S1) and the charge and hydrophobicity distribution for all included sequences plotted and structured according to the phylogenetic tree generated from the MSA (supplementary S2).

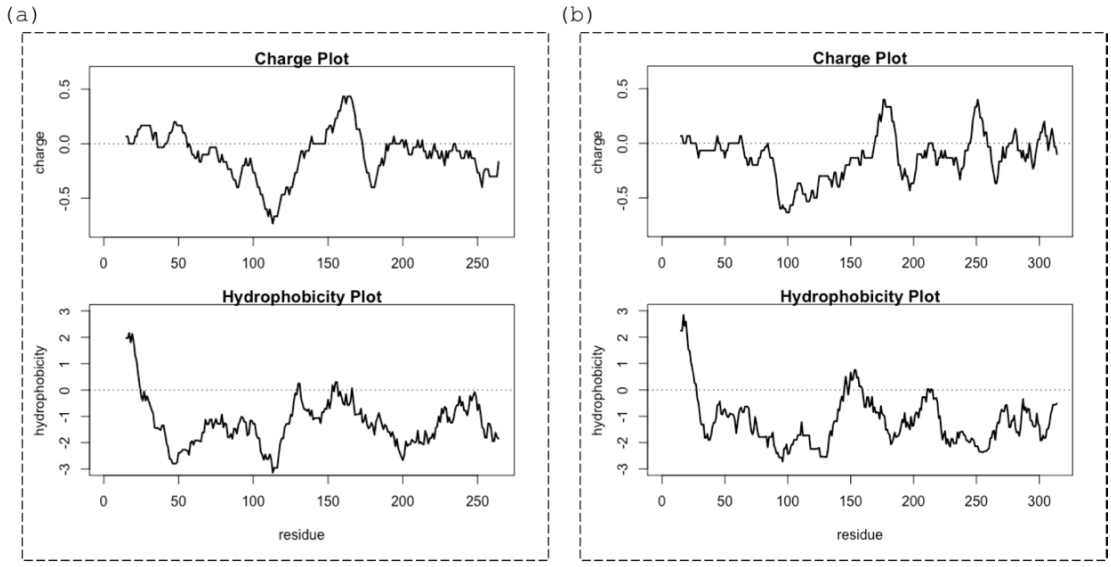

Figure 3.1. Charge and hydrophobicity plots of qOPN (a) and hOPN (b). The residue number is plotted on the x-axis with the corresponding hydrophobicity/charge values on the y-axis. Hydrophobicity was calculated according to the scale of Kyte & Doolittle [84] and charge by setting charged residues to +1 or -1 respectively (for histidine a value of +0.5 was assumed). For each residue an average value from a window of size 15 was calculated.

Figure 3.1 shows the charge and hydrophobicity distribution plots for quail (a) and human OPN (b). The range of the hydrophobicity plot is -4.5 to 4.5, with high values corresponding to hydrophobic regions. In the charge distribution plots, very positive or negative values correspond to respectively highly charged regions within a range of -1 to 1. From this analysis an overall good conservation for both charge and hydrophobicity distribution can be observed, especially when considering the low sequence conservation of OPN.

The high hydrophobicity values at the beginning of the sequence correspond to the signal peptide of OPN. The well conserved region surrounding the RGD motif, residues 138-140 in qOPN and 159-161 in hOPN, is of particular interest. In all analysed sequences, an increase of negatively charged residues before the integrin binding motif, followed by an increase in positively charged residues after the motif can be observed. Finally, there is another increase in negative charges close to position 180 in qOPN and 200 in hOPN. This charge distribution already indicates a certain compaction potential of this area driven by electrostatic interactions. The region directly preceding the RGD motif has a conserved increase in the amount of hydrophobic residues and prolines.

Analysis of Long-Range Contacts in OPN using PRE

PRE rates for ten qOPN (S54C, E77C, S108C, R132C, A161C, D174C, S188C, S206C, S228C, S247C) and nine hOPN (G25C, Q50C, D90C, D130C, T185C, D210C, S239C, S267C, S311C) spin label positions were calculated from the para- and diamagnetic measurements of each mutant. The PRE profiles can be viewed in supplementary S3 for qOPN and S4 for hOPN. High PRE rates correspond to an effect of the spin label on the affected areas, indicating spacial proximity.

In all profiles, a long-range effect of the paramagnetic label on different regions, not only the area close to the label on the primary sequence of the protein, can be seen. This already suggests a behaviour deviating from a theoretical random coil, which would generate an even distribution of PRE rates for residues not surrounding the spin label site. This can be seen most prominently for labels placed close to the termini, where distinct areas show an increase in PRE rates. Labels positioned closer to the centre of the proteins generally affect the entire surrounding region, indicating a compaction of these areas.

Figure 3.2 shows example profiles generated from the PRE rates of the mutant E77C for qOPN (a) and D90C for hOPN (b). The most relevant regions, where the difference in transverse relaxation rates between the paramagnetic and diamagnetic form is large are indicated for both profiles. Residues in region A are strongly affected by the label due to their proximity on the primary sequence, while effects for residues in regions B and C can only be explained by spacial proximity to the label through long-range contacts occurring in the protein.

From the PRE profiles of all mutants, a correlation matrix was calculated for both proteins. Positive correlations (red) are obtained when two residues are affected in a similar way by the paramagnetic labels due to concerted structural elements, while anti-correlation (blue) between two regions indicates a low structural relation within the conformational ensemble of the protein.

The correlation map for qOPN (figure 3.3a) shows a strong overall compaction of the protein, indicated by the high number of strongly correlated residues. Both the N- and C-Terminus show a strong local compaction as well as the entire middle region of the protein (residues 75 to 210). In both terminal regions, there are areas weakly correlated with the centre of the protein, while there is a negative correlation between the two segments.

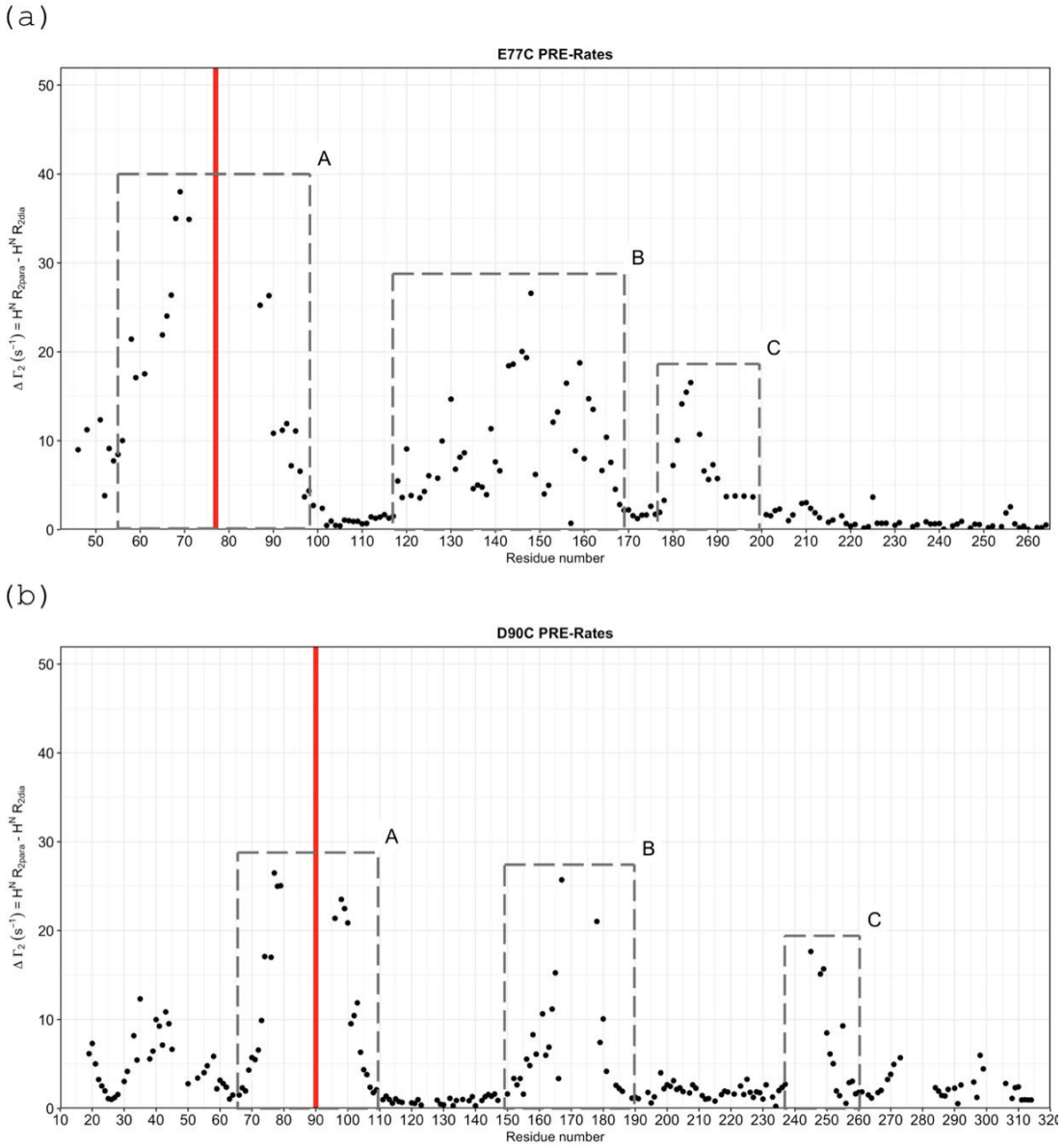

Figure 3.2. Examples for PRE profiles of qOPN-E77C (a) and hOPN-D90C (b). The PRE rates Γ_2 were calculated as the difference between the transverse relaxation rates (R_2) for the para- and diamagnetic state of each residue. The position of the paramagnetic label is marked by the red bar. Strongly affected regions are indicated by boxes A, B and C.

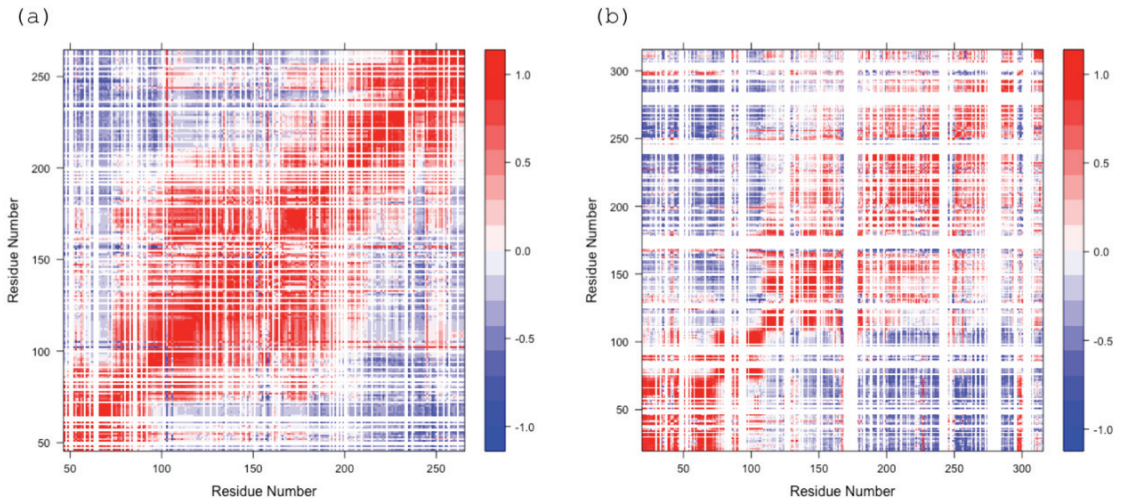

Figure 3.3. Correlation matrices of qOPN (a) and hOPN (b). The correlation coefficient was calculated for each residue-pair taking all spin label positions (10 for qOPN and 9 for hOPN) into account. Correlated regions are coloured red, while anti-correlated regions are blue.

Looking at the correlation matrix of hOPN (figure 3.3b), especially the N-Terminal segment shows a strong correlation, suggesting a highly compacted region. Similar to qOPN, there is an overall compacted segment of the protein, although in the case of hOPN a larger region (residues 110 to 314), which is more shifted towards the C-Terminus is affected with a less pronounced correlation.

Analysis of the OPN Interaction Network

Phosphorylation of hOPN

The phosphorylation of hOPN was confirmed via Mass Spectrometry (MS) Analysis and NMR, additionally the phosphorylation sites were identified using both techniques. A comparison of the phosphorylation sites found in the literature with the NMR and MS results from our sample can be viewed in table 3.1. Overall, a good agreement between the different samples could be observed.

An overlap of the hOPN spectra before and after phosphorylation in figure 3.4 shows significant shifts for several residues. Especially the shift of serine residues upon phosphorylation can be clearly recognised. A closer look at the serine region is provided in figure 3.4, indicating the chemical shift change of serine at position 26 upon phosphorylation as an example.

Residue	Milk	HepG2	MS	NMR	Residue	Milk	HepG2	MS	NMR
S24	+		+		S195	+	+	+	+
S26	+	+	+	+	S215	+	+	+	+
S27	+	+	+	+	S219	+	+	+	+
S62	+	+	+	+	S224	+	+	+	+
S63	+	+	+	+	Y225		+		
T66	+				T227			+	
S78			*	+	S228	+	+	+	+
S76	+		*		S234	+	+	+	+
S81	+		*	+	T237		+		+
S99	+		*	+	S239		+	+	*
S102	+		*		S243		+		*
S105	+		*	+	S254	+	+	+	+
S108	+		*	+	S258		+	+	+
S117	+		*	+	S263	+	+	+	+
S120	+		*	+	S267	+	+		
S123	+		*	+	S270		+	+	+
S126	+		*	+	S275	+	+	+	+
S129	+		*	+	S280	+	+	+	+
S162			+		S291	+	+	+	*
Y165			+		S303	+	+	+	*
S169			+		S308	+	+	+	+
T185	+		+	+	S310	+	+	+	+
T190		+	+		S311	+	+	+	(+)
S191	+	+	+	+					

Table 3.1. Comparison of phosphorylation sites found in human milk OPN (Milk) [58], OPN phosphorylated by Fam20c in HepG2 cells (HepG2) [61] and our sample analysed by MS and NMR analysis (MS and NMR respectively). A "+" indicates a phosphorylation, residues not covered by MS or not assigned with NMR are indicated by an asterisk. For S311 a notable upfield shift in the proton dimension indicates a modification of this residue, although generally a downfield shift is expected for phosphorylation.

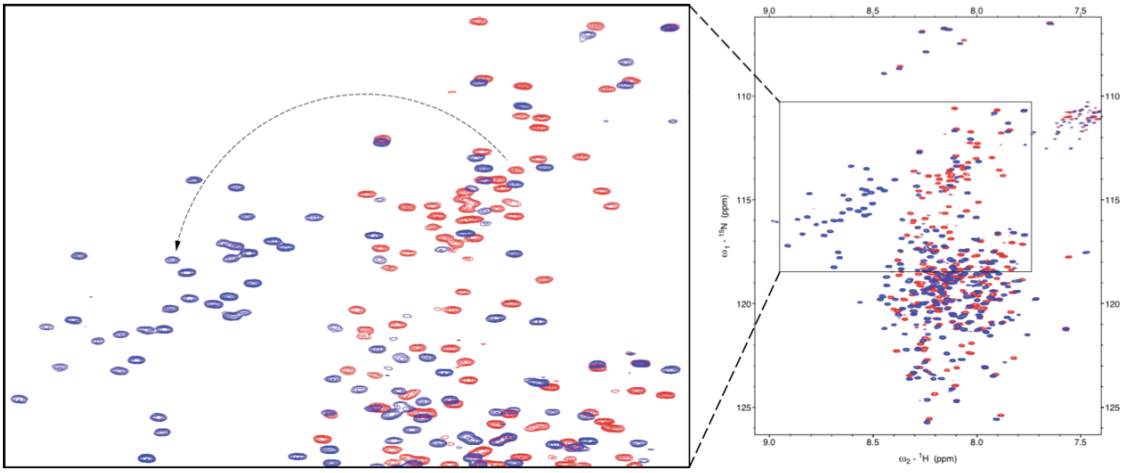

Figure 3.4. The figure shows an overlap of the spectra from hOPN before (red) and after (blue) phosphorylation. The most prominent shifts are seen for the serine region (zoomed region on the left) due to the phosphorylation of these residues. The arrow shows the shift of S26 due to phosphorylation of the residue as an example.

Interaction of hOPN with CD44stem

The binding of OPN to the stem of CD44 was confirmed via NMR, both from the side of hOPN and CD44stem. An overlap of the hOPN spectra before and after the addition of CD44stem can be seen in supplementary S5. A change in intensity for several residues could be observed, indicating binding of hOPN to CD44stem; the same was observed for CD44stem upon the addition of hOPN.

Interaction of hOPN with Hyaluronic Acid and Heparin

The binding of hyaluronic acid and heparin to hOPN before and after phosphorylation was analysed by adding increasing molar ratios of the biomolecules to an hOPN sample; all resulting spectra of the complete titrations can be found in the supplementary material (supplementary S6-9).

No significant shifts were observed for the titration of hyaluronic acid to hOPN before phosphorylation, even at high concentrations, shown in supplementary S6. Upon titration of hyaluronic acid to phosphorylated hOPN (supplementary S8), small shifts were observed for some phosphorylated residues.

For heparin, binding to hOPN was confirmed before and after phosphorylation (supplementary S7 and S9). Higher concentrations of heparin were needed to observe significant chemical shift changes in the phosphorylated sample and overall less changes were observed. Figure 3.5 demonstrates this effect showing the chemical shift changes for a residue upon heparin titration for the non-phosphorylated (a) and phosphorylated (b) hOPN.

The chemical shift perturbations (CSP) caused by the titration of heparin to hOPN are shown in figure 3.6. In the non-phosphorylated sample, especially residues in the regions 130-190 (with a dip around 160) and 230-270 show high values, demonstrating a strong change in the chemical shift.

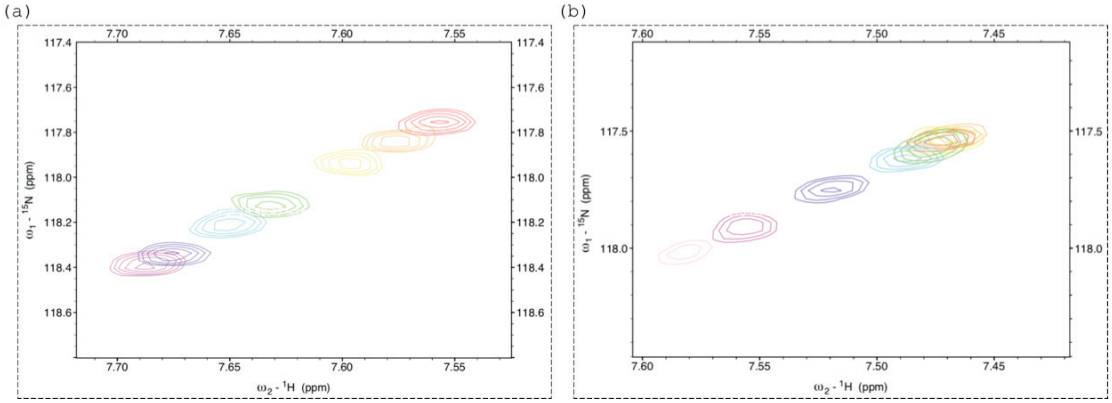

Figure 3.5. Chemical shifts of residue I179 during titration of heparin for non-phosphorylated OPN (a) and phosphorylated OPN (b). The peak before addition of heparin is shown in red. Titrations were carried out in molar ratios of 10:1 (orange), 5:1 (yellow), 2:1 (green), 1:1 (cyan), 1:2 (blue) and 1:5 (magenta) another titration step of 1:10 (pink) was added to the phosphorylated sample.

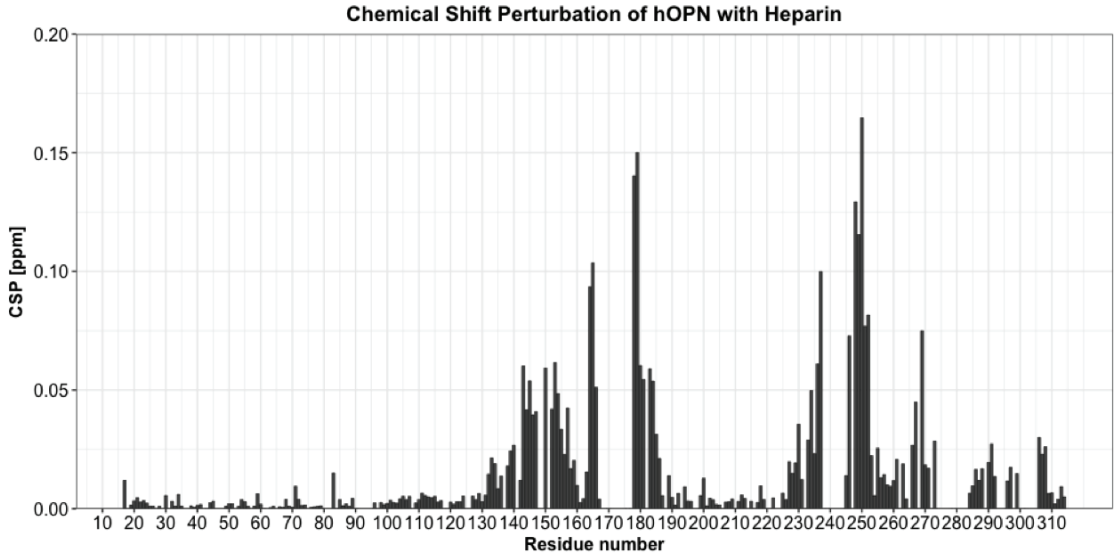

Figure 3.6. Chemical Shift Perturbations through the addition of heparin to hOPN.

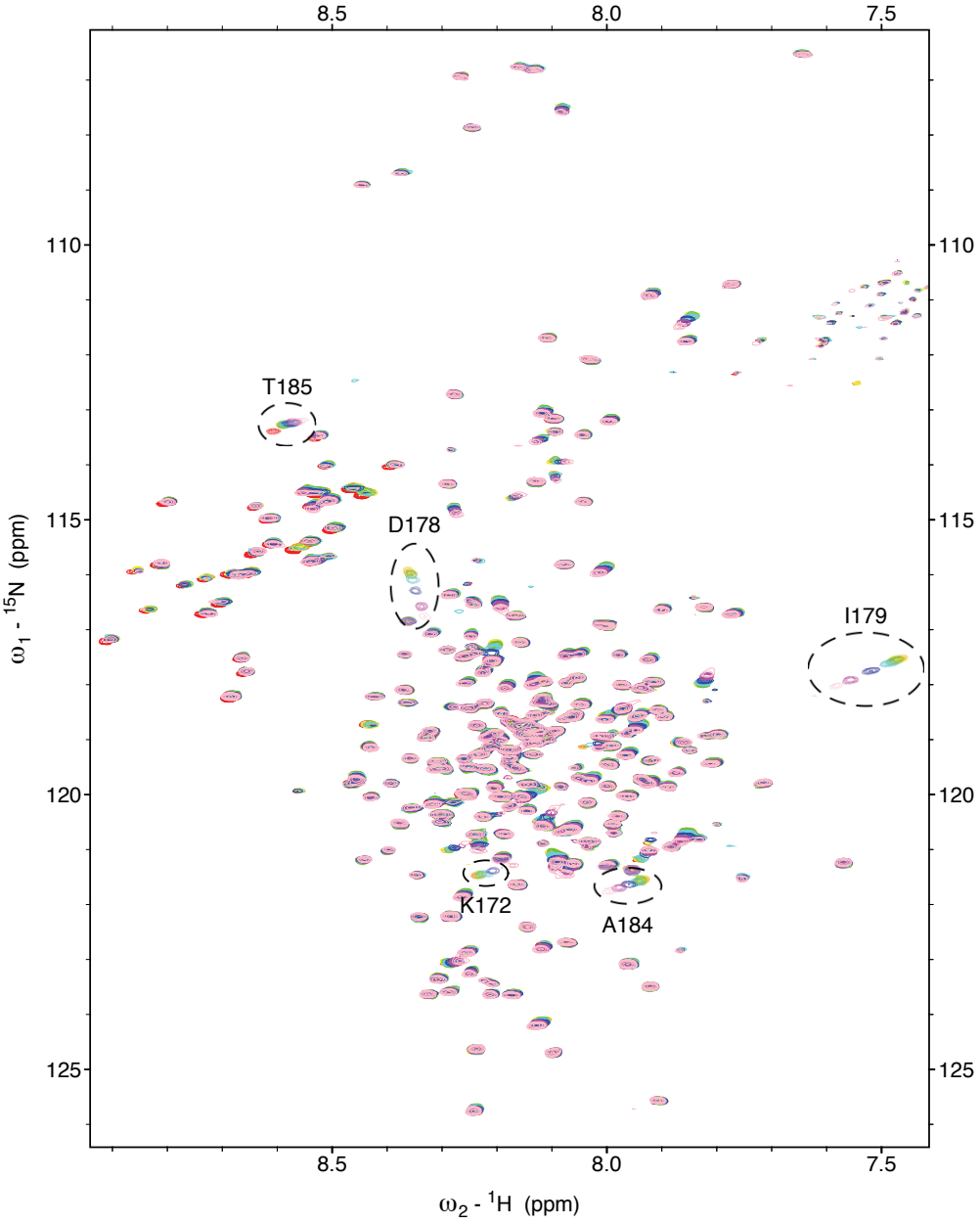

Figure 3.7. Titration of heparin with phosphorylated hOPN. Residues with a strong chemical shift change during the titration are indicated.

In figure 3.7 the spectra of the heparin titration with the phosphorylated hOPN is shown; residues with a large shift are indicated. Changes in the chemical shift are primarily observed for residues in the region 170-185, small shifts could be shown for some residues in the positively charged region at ca. residue 250.

Interaction of hOPN with the Hyaluronic Acid Binding Domain of CD44

An analysis of hyaluronic acid binding to the HABD of CD44 was conducted by recording 1D spectra of the HABD with hOPN and hyaluronic acid. The spectrum of HABD shows a prototypical chemical shift dispersion of a folded protein. The dashed box in figure 3.8 shows that signals from the the HABD disappear upon addition of hyaluronic acid, indicating an interaction. 2D spectra recorded from hOPN in the presence of the HABD with and without hyaluronic acid showed no interaction between the two proteins.

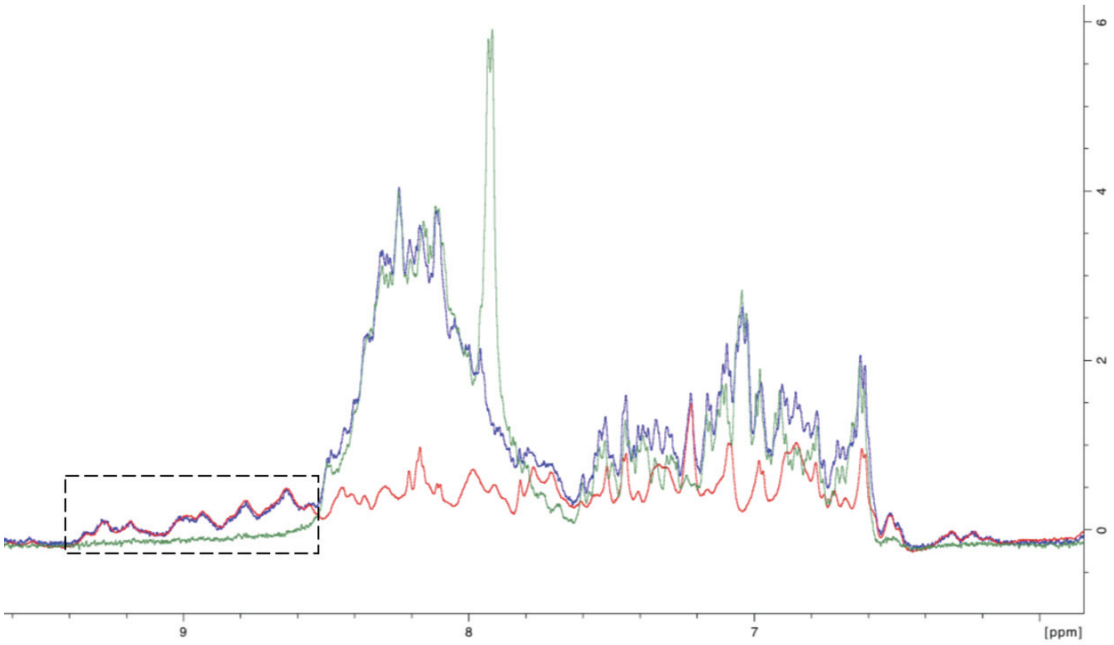

Figure 3.8. An overlap of the 1D-Spectra of HABD. The spectrum of HABD is shown in red, the HABD with hOPN is shown in blue and the HABD with hOPN and hyaluronic acid in green. The signal at ca. 7.8ppm is from the hyaluronic acid. The region indicated by the dashed box shows that upon addition of hyaluronic acid, signals from the HABD are significantly reduced in intensity.

Discussion

The ability to dynamically adopt extended as well as compact conformations is an important feature of IDPs, enabling their versatile biological functions, which makes it worthwhile to study what drives the transitions between these states and how they are regulated. From the long-range effects of the different spin label positions and the resulting correlation maps, it can be concluded that both hOPN and qOPN adopt compact conformations.

For both quail and human OPN, a significant effect on the PRE rates close to the RGD motif could be observed. The lower overall PRE rates and correlations found for hOPN indicate the compact conformations are either less populated or they are less spatially packed than in qOPN. This suggests qOPN and hOPN have different structural pre-formation strategies. In hOPN, a strongly affected region surrounding residue 250 shows an increase in positively charged residues, which probably leads to additional tertiary contacts due to electrostatic interactions. While this stretch of positive residues is missing in qOPN, a region strongly affected by the spin labels corresponds to a tryptophane (W184) followed by a proline. This is an intriguing motif, as aromatics (especially tryptophane) close to a proline stabilise the cis-conformation [85], which results in a turn-like structure compared to the locally extended conformations of trans-isoforms [86].

Previous studies have shown that the presence of both urea and NaCl is required to unfold qOPN [74], indicating an intricate interplay of several compaction mechanisms. Considering the low sequence identity of OPN between species, the overall biophysical properties are well conserved, especially the distribution of charged and hydrophobic residues surrounding the RGD motif responsible for the binding of OPN to integrins. The opposing charges before and after the motif already indicate tertiary contacts between these regions as confirmed by the PRE experiments. The high number of prolines and hydrophobic residues directly preceding the RGD motif is also well conserved, suggesting an important role of prolines in modulating the complex network of intramolecular contacts that lead to compaction.

OPN is highly phosphorylated in-vivo, which affects the binding to several partners and could have important implications for the compaction of OPN through the addition of negative charges. Most likely the extensive phosphorylation of OPN will lead to an overall expansion, preventing the formation of strongly compacted regions. The phosphorylation sites of hOPN were mapped using MS and NMR. With NMR, additional sites could be identified not covered by MS, while MS was able to map lowly populated phosphorylation sites not detected by NMR. Some of the observed phosphorylation sites do not follow the canonical Fam20c motif (Ser-X-Glu/pSer), suggesting a certain degree of promiscuity of this kinase. The overall good agreement between the phosphorylation sites found in the literature compared to the sample generated by in-vitro phosphorylation indicates the sample can be used to further investigate the effect of phosphorylation on the conformation and interactions of OPN. By MS three low populated non-canonical sites close to the RGD motif (S162, Y165, S169) were identified. These sites could have important implications on modulating the interaction between OPN and integrins. Thrombin cleavage is important for the efficient binding of OPN to several integrins, which makes the phosphorylation found at the thrombin cleavage site (S169) especially compelling, as it has been shown that phosphorylation was able to inhibit thrombin cleavage of tau [87]. These findings will be further investigated due to the potential biological implication of OPN phosphorylation by Fam20c.

The effect of phosphorylation on the interaction of hOPN to several binding partners was analysed. Like in qOPN, the binding of heparin to hOPN was confirmed by NMR. Several chemical shift changes were observed and the binding sites mapped by analysing the CSP of all residues upon addition of heparin. Heparin binds to the positively charged regions (surrounding positions 175 and 250) due to electrostatic attraction of the negatively charged biomolecule. The second binding region was not observed in qOPN, which does not have this positive patch. The chemical shift changes for residues preceding the RGD motif could be due to the expansion of the compact state upon heparin binding as has been suggested for heparin binding to qOPN [88]. For the binding of heparin to phosphorylated hOPN less shifts were observed, primarily in the positively charged area following the RGD motif. No shifts were observed for the negatively charged region, which could indicate the compact conformation is not present in the phosphorylated sample and therefore is not modulated upon the binding of heparin.

The interaction studies showed hOPN does not bind to hyaluronic acid and the HABD of CD44, although binding to both heparin and the flexible stem of CD44 could be observed. The binding of hOPN to heparin could have important implications for the interaction with CD44, as the splicing variant 3 of CD44 contains a heparan sulphate modification, which is structurally related to heparin (supplementary S10). Future experiments will be conducted analysing the binding of hOPN to heparan sulphate to further explore this hypothesis. The interaction possibilities of hOPN and CD44 were delimited: hOPN can only bind to CD44 through the stem part and/or the heparan sulphate PTM, which could be modulated by phosphorylation.

Conclusion

This thesis presents a characterisation of the compact states of both qOPN and hOPN as well as an analysis of the complex interaction network of hOPN using NMR techniques. It has been demonstrated that for IDPs, which usually have a low sequence identity across species, it is particularly interesting to investigate conserved similarities in biophysical properties.

This work provides a more complete picture of the compact state of qOPN through PRE measurements of additional spin label positions. It also provides the first experiments focussing on the compact state of hOPN using the same approach. Although certain sequence features are well conserved between the species, significant changes between human and quail OPN compaction were identified. This indicates additional sidechain and backbone interactions, such as the one occurring in aromatic-proline motifs, may play a role in the subtle regulation of their conformational ensemble.

In addition to protein sequence properties, PTMs, such as phosphorylation, are able to modulate long-range contacts in IDPs. The in-vitro phosphorylation of hOPN by Fam20c was carried out. The generation of a soluble, highly phosphorylated and pure sample allows atomic resolution studies by NMR. Moreover, the results gained from the binding studies pave the way to further characterise the interaction of OPN with CD44 and biomolecules of the extracellular matrix, like heparin and hyaluronic acid.

Bibliography

[1] Christopher J. Oldfield and A. Keith Dunker. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry*, 83(1):553–584, 2014.

[2] Ruma Banerjee. Introduction to the thematic minireview series on intrinsically disordered proteins. *The Journal of Biological Chemistry*, 291(13):6679–6680, 03 2016.

[3] Zsuzsanna Dosztányi, Jake Chen, A. Keith Dunker, István Simon, and Peter Tompa. Disorder and sequence repeats in hub proteins and their implications for network evolution. *Journal of Proteome Research*, 5(11):2985–2995, 11 2006.

[4] Amelie Stein, Roland A. Pache, Pau Bernadó, Miquel Pons, and Patrick Aloy. Dynamic interactions of proteins in complex networks: a more structured view. *The FEBS Journal*, 276(19):5390–5405, 2009.

[5] Veronika Csizmok, Ariele Viacava Follis, Richard W. Kriwacki, and Julie D. Forman-Kay. Dynamic protein interaction networks and new structural paradigms in signaling. *Chemical Reviews*, 116(11):6424–6462, 06 2016.

[6] Zhenling Peng, Jing Yan, Xiao Fan, Marcin J. Mizianty, Bin Xue, Kui Wang, Gang Hu, Vladimir N. Uversky, and Lukasz Kurgan. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cellular and Molecular Life Sciences*, 72(1):137–151, 2015.

[7] A. Keith Dunker, Sarah E. Bondos, Fei Huang, and Christopher J. Oldfield. Intrinsically disordered proteins and multicellular organisms. *Seminars in Cell & Developmental Biology*, 37:44–55, 2015.

[8] Peter E. Wright and H. Jane Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews. Molecular cell biology*, 16(1):18–29, 01 2015.

[9] Jörg Gsponer, Matthias E. Futschik, Sarah A. Teichmann, and M. Madan Babu. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science (New York, N.Y.)*, 322(5906):1365–1368, 11 2008.

[10] M. Madan Babu, Robin van der Lee, Natalia Sanchez de Groot, and Jörg Gsponer. Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology*, 21(3):432–440, 2011.

[11] Tanya Vavouri, Jennifer I. Semple, Rosa Garcia-Verdugo, and Ben Lehner. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell*, 138(1):198–208, 07 2009.

[12] Vladimir N. Uversky, Christopher J. Oldfield, and A. Keith Dunker. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annual Review of Biophysics*, 37(1):215–246, 07 2008.

[13] Robert Konrat. IDPs: less disordered and more ordered than expected. *Biophysical Journal*, 109(7):1309–1311, 10 2015.

[14] H. Jane Dyson and Peter E. Wright. Intrinsically unstructured proteins and their functions. *Nature Reviews. Molecular Cell Biology*, 6, 03 2005.

[15] Julie D. Forman-Kay and Tanja Mittag. From sequence and forces to structure, function and evolution of intrinsically disordered proteins. *Structure (London, England : 1993)*, 21(9):1492–1499, 09 2013.

[16] Joseph A. Marsh and Julie D. Forman-Kay. Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical Journal*, 98(10):2383–2390, 05 2010.

[17] Shaolong Zhu, Agnesa Shala, Alexandr Bezginov, Adnan Sljoka, Gerald Audette, and Derek J. Wilson. Hyperphosphorylation of intrinsically disordered tau protein induces an amyloidogenic shift in its conformational ensemble. *PLoS ONE*, 10(3), 2015.

[18] Dominick Lemas, Panagiotis Lekkas, Bryan A. Ballif, and Jim O. Vigoreaux. Intrinsic disorder and multiple phosphorylations constrain the evolution of the flightin n-terminal region. *Journal of proteomics*, 135:191–200, 03 2016.

[19] Alaji Bah and Julie D. Forman-Kay. Modulation of intrinsically disordered protein function by post-translational modifications. *The Journal of Biological Chemistry*, 291(13):6696–6705, 03 2016.

[20] Marija Buljan, Guilhem Chalancon, Sebastian Eustermann, Gunter P. Wagner, Monika Fuxreiter, Alex Bateman, and M. Madan Babu. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Molecular Cell*, 46(6):871–883, 06 2012.

[21] Marija Buljan, Guilhem Chalancon, A. Keith Dunker, Alex Bateman, S. Balaji, Monika Fuxreiter, and M. Madan Babu. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Current Opinion in Structural Biology*, 23(3):443–450, 2013.

[22] J. Sodek, B. Ganss, and M.D. McKee. Osteopontin. *Critical Reviews in Oral Biology & Medicine*, 11(3):279–303, 2000.

[23] Katherine A. Staines, Vicky E. MacRae, and Colin Farquharson. The importance of the SIBLING family of proteins on skeletal mineralisation and bone remodelling. *Journal of Endocrinology*, 214(3):241–255, 09 2012.

[24] Akeila Bellahcène, Vincent Castronovo, Kalu U. E. Ogbureke, Larry W. Fisher, and Neal S. Fedarko. Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): Multifunctional proteins in cancer. *Nature reviews. Cancer*, 8(3):212–226, 03 2008.

[25] Donald R. Senger, Carole A. Perruzzi, Ageliki Papadopoulos, and Daniel G. Tenen. Purification of a human milk protein closely similar to tumor-secreted phosphoproteins and osteopontin. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 996(1):43–48, 1989.

[26] Patrizia Lanteri. *Clinical Chemistry and Laboratory Medicine*, volume 50, chapter Stability of osteopontin in plasma and serum, page 1979. 2012.

[27] Brian M. Nolen, Lidiya S. Orlichenko, Adele Marrangoni, Liudmila Velikokhatnaya, Denise Prosser, William E. Grizzle, Kevin Ho, Frank J. Jenkins, Dana H. Bovbjerg, and Anna E. Lokshin. An Extensive Targeted Proteomic Analysis of Disease-Related Protein Biomarkers in Urine from Healthy Donors. *PLoS ONE*, 8(5), 05 2013.

[28] D. T. Denhardt and X. Guo. Osteopontin: a protein with diverse functions. *The FASEB Journal*, 7(15):1475–1482, 1993.

[29] Florian Kahles, Hannes M. Findeisen, and Dennis Bruemmer. Osteopontin: A novel regulator at the cross roads of inflammation, obesity and diabetes. *Molecular Metabolism*, 3(4):384–393, 07 2014.

[30] S. R. Rittling and R. Singh. Osteopontin in immune-mediated diseases. *Journal of Dental Research*, 94(12):1638–1645, 12 2015.

[31] Nausicaa Clemente, Davide Raineri, Giuseppe Cappellano, Elena Boggio, Francesco Favero, Maria Felicia Soluri, Chiara Dianzani, Cristoforo Comi, Umberto Dianzani, and Annalisa Chiocchetti. Osteopontin bridging innate and adaptive immunity in autoimmune diseases. *Journal of Immunology Research*, 2016(7675437), 2016.

[32] S. R. Rittling and A. F. Chambers. Role of osteopontin in tumour progression. *British Journal of Cancer*, 90(10):1877–1881, 05 2004.

[33] Ran Wei, Janet Pik Ching Wong, and Hang Fai Kwok. Osteopontin – a promising biomarker for cancer therapy. *Journal of Cancer*, 8(12):2173–2183, 2017.

[34] L.W. Fisher, D.A. Torchia, B. Fohr, M.F. Young, and N.S. Fedarko. Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochemical and Biophysical Research Communications*, 280(2):460–465, 2001.

[35] Adele L. Boskey. Osteopontin and related phosphorylated sialoproteins: Effects on mineralization. *Annals of the New York Academy of Sciences*, 760(1):249–256, 1995.

[36] Graeme K. Hunter. Role of osteopontin in modulation of hydroxyapatite formation. *Calcified Tissue International*, 93(4):348–354, 2013.

[37] Luigi Mario Castello, Davide Raineri, Livia Salmi, Nausicaa Clemente, Rosanna Vaschetto, Marco Quaglia, Massimiliano Garzaro, Sergio Gentilli, Paolo Navalesi, Vincenzo Cantaluppi, Umberto Dianzani, Anna Aspesi, and Annalisa Chiocchetti. Osteopontin at the crossroads of inflammation and tumor progression. *Mediators of Inflammation*, (4049098), 2017.

[38] Anthony O'Regan and Jeffrey S. Berman. Osteopontin: a key cytokine in cell-mediated and granulomatous inflammation. *International Journal of Experimental Pathology*, 81(6):373–390, 12 2000.

[39] Yasuyuki Yokosaki, Kumi Tanaka, Fumiko Higashikawa, Keisuke Yamashita, and Akira Eboshida. Distinct structural requirements for binding of the integrins $\alpha\beta6$, $\alpha v\beta3$, $\alpha v\beta5$, $\alpha5\beta1$ and $\alpha9\beta1$ to osteopontin. *Matrix Biology*, 24(6):418–427, 2005.

[40] Georg F. Weber, Samy Ashkar, Melvin J. Glimcher, and Harvey Cantor. Receptor-ligand interaction between CD44 and osteopontin (eta-1). *Science*, 271(5248):509–512, 1996.

[41] Georg F. Weber. The metastasis gene osteopontin: a candidate target for cancer therapy. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1552(2):61–85, 2001.

[42] Helmut Ponta, Larry Sherman, and Peter A. Herrlich. CD44: From adhesion molecules to signalling regulators. *Nature Reviews Molecular Cell Biology*, 4, 01 2003.

[43] David T. Denhardt, Masaki Noda, Anthony W. O'Regan, Dubravko Pavlin, and Jeffrey S. Berman. Osteopontin as a means to cope with environmental insults: regulation of inflammation, tissue remodeling, and cell survival. *Journal of Clinical Investigation*, 107(9):1055–1061, 05 2001.

[44] Yohko U. Katagiri, Jonathan Sleeman, Hideki Fujii, Peter Herrlich, Hiroshi Hotta, Kumiko Tanaka, Shunsuke Chikuma, Hideo Yagita, Ko Okumura, Masaaki Murakami, Ikuo Saiki, Ann F. Chambers, and Toshimitsu Uede. CD44 variants but not CD44s cooperate with β1-containing integrins to permit cells to bind to osteopontin independently of arginine-glycine-aspartic acid, thereby stimulating cell motility and chemotaxis. *Cancer Research*, 59(1):219–226, 1999.

[45] Margot Zöller. Cd44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nature Reviews Cancer*, 11, 03 2011.

[46] Renu Agnihotri, Howard C. Crawford, Hirotaka Haro, Lynn M. Matrisian, Matthew C. Havrda, and Lucy Liaw. Osteopontin, a novel substrate for matrix metalloproteinase-3 (stromelysin-1) and matrix metalloproteinase-7 (matrilysin). *Journal of Biological Chemistry*, 276(30):28261–28267, 2001.

[47] V. Takafuji, M. Forgues, E. Unsworth, P. Goldsmith, and X. W. Wang. An osteopontin fragment is essential for tumor cell invasion in hepatocellular carcinoma. *Oncogene*, 26, 04 2007.

[48] Yoshiki Yamaguchi, Shinya Hanashima, Hirokazu Yagi, Yutaka Takahashi, Hiroaki Sasakawa, Eiji Kurimoto, Takeshi Iguchi, Shigeyuki Kon, Toshimitsu Uede, and Koichi Kato. NMR characterization of intramolecular interaction of osteopontin, an intrinsically disordered protein with cryptic integrin-binding motifs. *Biochemical and Biophysical Research Communications*, 393(3):487–491, 2010.

[49] Yasuyuki Yokosaki, Nariaki Matsuura, Tomohiro Sasaki, Isao Murakami, Holm Schneider, Shigeki Higashiyama, Yoshiki Saitoh, Michio Yamakido, Yasuyuki Taooka, and Dean Sheppard. The integrin α9β1 binds to a novel recognition sequence (SVVYGLR) in the thrombin-cleaved amino-terminal fragment of osteopontin. *Journal of Biological Chemistry*, 274(51):36328–36334, 12 1999.

[50] Mari L. Shinohara, Hye-Jung Kim, June-Ho Kim, Virgilio A. Garcia, and Harvey Cantor. Alternative translation of osteopontin generates intracellular and secreted isoforms that mediate distinct biological activities in dendritic cells. *Proceedings of the National Academy of Sciences of the United States of America*, 105(20):7235–7239, 05 2008.

[51] Makoto Inoue and Mari L. Shinohara. Intracellular osteopontin (iOPN) and immunity. *Immunologic research*, 49(1-3):160–172, 04 2011.

[52] E. R. Gimba and T. M. Tilli. Human osteopontin splicing isoforms: Known roles, potential clinical applications and activated signaling pathways. *Cancer Letters*, 331(1):11–17, 07 2013.

[53] A. Gericke, C. Qin, L. Spevak, Y. Fujimoto, W. T. Butler, E. S. Sørensen, and A. L. Boskey. Importance of phosphorylation for osteopontin regulation of biomineralization. *Calcified tissue international*, 77(1):45–54, 07 2005.

[54] Weber Georg F., Zawaideh Samer, Hikita Sherry, Kumar Vikram A., Cantor Harvey, and Ashkar Samy. Phosphorylation-dependent interaction of osteopontin with its receptors regulates macrophage migration and activation. *Journal of Leukocyte Biology*, 72(4):752–761, 07 2002.

[55] Mandana Keykhosravani, Amanda Doherty-Kirby, Cunjie Zhang, Dyanne Brewer, Harvey A. Goldberg, Graeme K. Hunter, and Gilles Lajoie. Comprehensive identification of post-translational modifications of rat bone osteopontin by mass spectrometry. *Biochemistry*, 44(18):6990–7003, 05 2005.

[56] Brian Christensen, Christian C. Kazanecki, Torben E. Petersen, Susan R. Rittling, David T. Denhardt, and Esben S. Sørensen. Cell type-specific post-translational modifications of mouse osteopontin are associated with different adhesive properties. *Journal of Biological Chemistry*, 282(27):19463–19472, 2007.

[57] E. S. Sørensen, P. Højrup, and T. E. Petersen. Posttranslational modifications of bovine osteopontin: identification of twenty-eight phosphorylation and three O-glycosylation sites. *Protein Science : A Publication of the Protein Society*, 4(10):2040–2049, 10 1995.

[58] Brian Christensen, Mette S. Nielsen, Kim F. Haselmann, Torben E. Petersen, and Esben S. Sørensen. 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. *Biochemical Journal*, 390(Pt 1):285–292, 08 2005.

[59] Brian Christensen, Torben E. Petersen, and Esben S. Sørensen. Post-translational modification and proteolytic processing of urinary osteopontin. *Biochemical Journal*, 411(1):53–61, 2008.

[60] Vincent S. Tagliabracci, James L. Engel, Jianzhong Wen, Sandra E. Wiley, Carolyn A. Worby, Lisa N. Kinch, Junyu Xiao, Nick V. Grishin, and Jack E. Dixon. Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science (New York, N.Y.)*, 336(6085):1150–1153, 06 2012.

[61] Vincent S Tagliabracci, Sandra E Wiley, Xiao Guo, Lisa N Kinch, Eric Durrant, Jianzhong Wen, Junyu Xiao, Jixin Cui, Kim B Nguyen, James L Engel, Joshua J Coon, Nick Grishin, Lorenzo A Pinna, David J Pagliarini, and Jack E Dixon. A single kinase generates the majority of the secreted phosphoproteome. *Cell*, 161(7):1619–1632, 06 2015.

[62] Hui Zhang, Qinyu Zhu, Jixin Cui, Yuxin Wang, Mark J Chen, Xing Guo, Vincent S Tagliabracci, Jack E Dixon, and Junyu Xiao. Structure and evolution of the Fam20 kinases. *Nature Communications*, 9:1218, 2018.

[63] M A Simpson, R Hsu, L S Keir, J Hao, G Sivapalan, L M Ernst, E H Zackai, L I Al-Gazali, G Hulskamp, H M Kingston, T E Prescott, A Ion, M A Patton, V Murday, A George, and A H Crosby. Mutations in Fam20c are associated with lethal osteosclerotic bone dysplasia (raine syndrome), highlighting a crucial molecule in bone development. *American Journal of Human Genetics*, 81(5):906–912, 11 2007.

[64] Simone Kosol, Sara Contreras-Martos, Cesyen Cedeño, and Peter Tompa. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules*, 18(9):10802–10828, 2013.

[65] Robert Konrat. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance*, 241:74–85, 2014.

[66] Mike P. Williamson. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73:1–16, 2013.

[67] Kragelj Jaka, Ozenne Valéry, Blackledge Martin, and Jensen Malene Ringkjøbing. Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *ChemPhysChem*, 14(13):3034–3045, 2013.

[68] G. Marius Clore. *Methods in Enzymology; Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions*, volume 564, chapter Practical Aspects of Paramagnetic Relaxation Enhancement in Biological Macromolecules, pages 485–497. Academic Press, 2015.

[69] G. Marius Clore and Junji Iwahara. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical Reviews*, 109(9):4108–4139, 09 2009.

[70] Peter H. J. Keizers and Marcellus Ubbink. Paramagnetic tagging for protein structure and dynamics analysis. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 58(1):88–96, 2011.

[71] Nicholas J. Anthis, Michaeleen Doucleff, and G. Marius Clore. Transient, sparsely-populated compact states of apo and calcium-loaded calmodulin probed by paramagnetic relaxation enhancement: interplay of conformational selection and induced fit. *Journal of the American Chemical Society*, 133(46):18966–18974, 11 2011.

[72] Junji Iwahara and G. Marius Clore. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, 440, 04 2006.

[73] G. Marius Clore. Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation. *Protein Science : A Publication of the Protein Society*, 20(2):229–246, 02 2011.

[74] Dennis Kurzbach, Gerald Platzer, Thomas C. Schwarz, Morkos A. Henen, Robert Konrat, and Dariush Hinderberger. Cooperative unfolding of compact conformations of the intrinsically disordered protein osteopontin. *Biochemistry*, 52(31):5167–5175, 08 2013.

[75] Gerald Platzer, Andreas Schedlbauer, Angela Chemelli, Przemyslaw Ozdowy, Nicolas Coudevylle, Renate Auer, Georg Kontaxis, Markus Hartl, Andrew J. Miles, B. A. Wallace, Otto Glatter, Klaus Bister, and Robert Konrat. The metastasis-associated extracellular matrix protein osteopontin forms transient structure in ligand interaction sites. *Biochemistry*, 50(27):6113–6124, 07 2011.

[76] Jordi Silvestre-Ryan, Carlos W. Bertoncini, Robert Bryn Fenwick, Santiago Esteban-Martin, and Xavier Salvatella. Average conformations determined from pre data provide high-resolution maps of transient tertiary interactions in disordered proteins. *Biophysical Journal*, 104(8):1740–1751, 08 2013.

[77] Kevin T. Dicker, Lisa A. Gurski, Swati Pradhan-Bhatt, Robert L. Witt, Mary C. Farach-Carson, and Xinqiao Jia. Hyaluronan: A simple polysaccharide with diverse biological functions. *Acta biomaterialia*, 10(4):1558–1570, 04 2014.

[78] Maria C. Z. Meneghetti, Ashley J. Hughes, Timothy R. Rudd, Helena B. Nader, Andrew K. Powell, Edwin A. Yates, and Marcelo A. Lima. Heparan sulfate and heparin interactions with proteins. *Journal of the Royal Society Interface*, 12(110), 09 2015.

[79] Frank Delaglio, Stephan Grzesiek, Geerten W. Vuister, Guang Zhu, John Pfeifer, and Ad Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293, 11 1995.

[80] T. D. Goddard and D. G. Kneller. *SPARKY 3*, 2008.

[81] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2016.

[82] Kazutaka Katoh, John Rozewicki, and Kazunori D. Yamada. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, page 108, 2017.

[83] Mira V. Han and Christian M. Zmasek. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, 10:356–356, 2009.

[84] Jack Kyte and Russell F. Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1):105–132, 1982.

[85] Neal J. Zondlo. Aromatic–proline interactions: Electronically tunable ch/πinteractions. *Accounts of Chemical Research*, 46(4):1039–1049, 04 2013.

[86] Francois-Xavier Theillet, Lajos Kalmar, Peter Tompa, Kyou-Hoon Han, Philipp Selenko, A Keith Dunker, Gary W Daughdrill, and Vladimir N Uversky. The alphabet of intrinsic disorder: I. act like a pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disordered Proteins*, 1(1):e24360, 2013.

[87] Tetsuaki Arai, Jian-Ping Guo, and Patrick L. McGeer. Proteolysis of non-phosphorylated and phosphorylated tau by thrombin. *Journal of Biological Chemistry*, 280(07):5145–5153, 2005.

[88] Andreas Beier, Thomas C. Schwarz, Dennis Kurzbach, Gerald Platzer, Francesca Tribuzio, and Robert Konrat. Modulation of correlated segment fluctuations in idps upon complex formation as an allosteric regulatory mechanism. *Journal of Molecular Biology*, 430(16):2439–2452, 2018.

[89] The UniProt Consortium. Uniprot: the universal protein knowledgebase. *Nucleic Acids Research*, 45(D1):D158–D169, 2017.

Supplementary Material

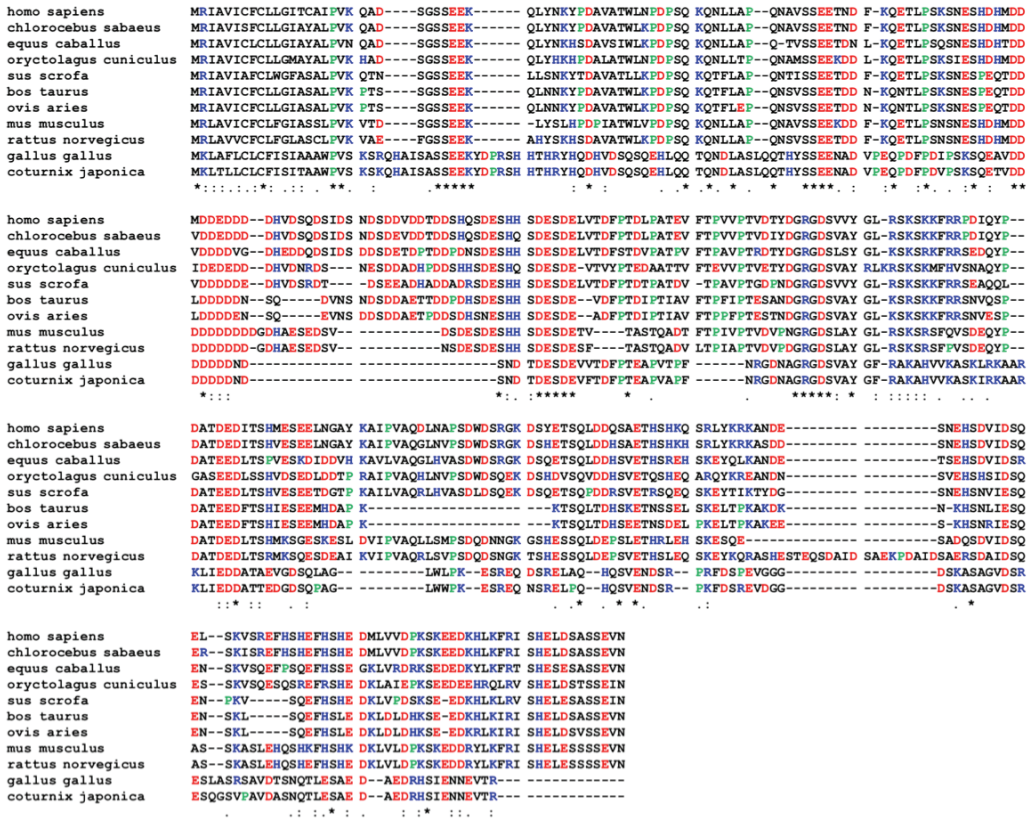

Figure S1. Multiple Sequence Alignment (MSA) of OPN generated from different sequences collected from UniProt [89] (P10451, A0A0D9QWB5, P10923, P08721, P31097, F7AYC1, P14287, P31096, Q9XSY9, P23498, Q9I832) using MAFFT [82]. Conserved residues with weakly and strongly similar properties are indicated by a "." and ":" respectively, identical residues by an asterisk. Prolines are coloured in green, positively charged residues in blue and negatively charged in red.

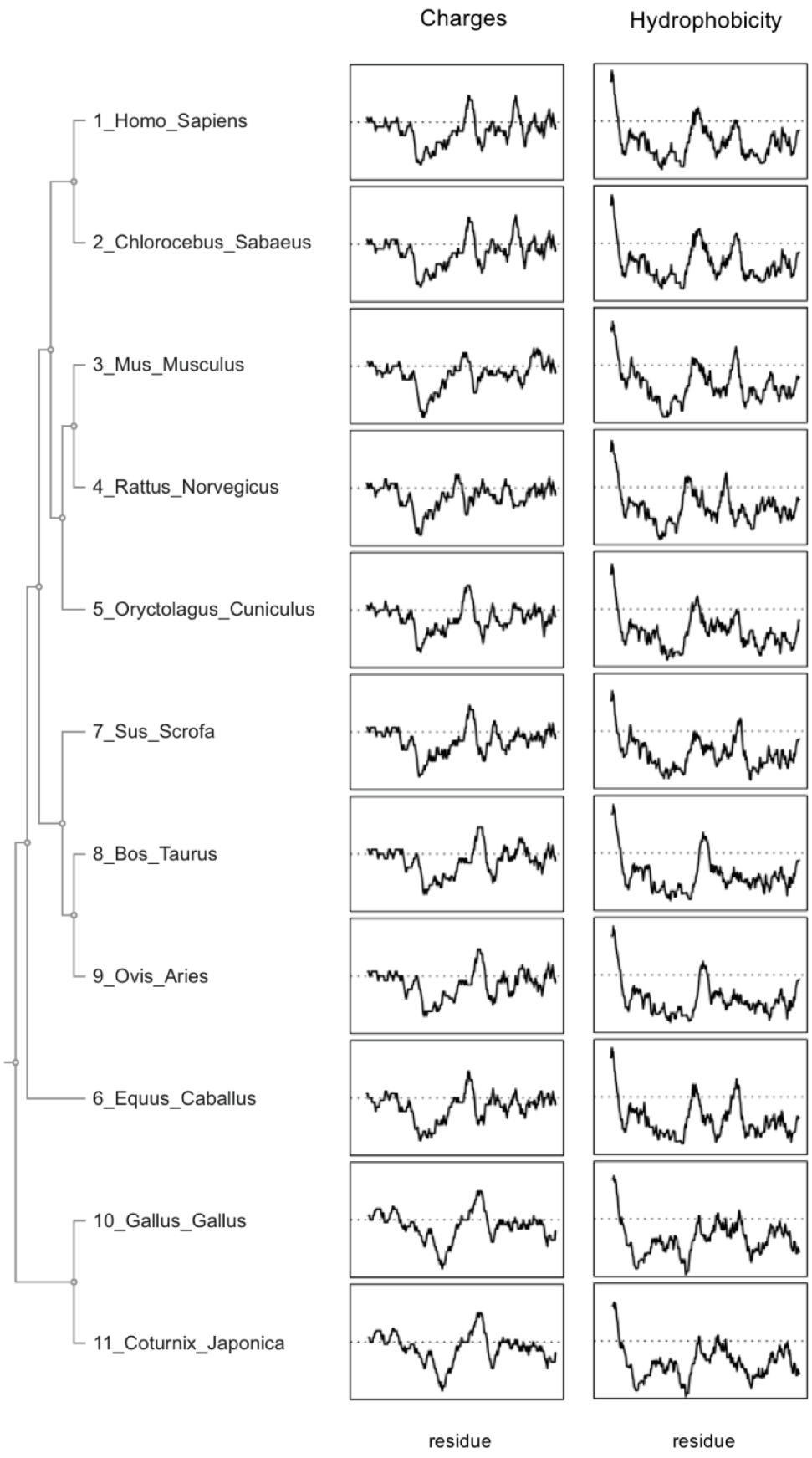

Figure S2. Charge and hydrophobicity plots of the OPN sequences used for the MSA. The tree is based on the MSA shown in figure S1. The dotted line indicates a value of zero.

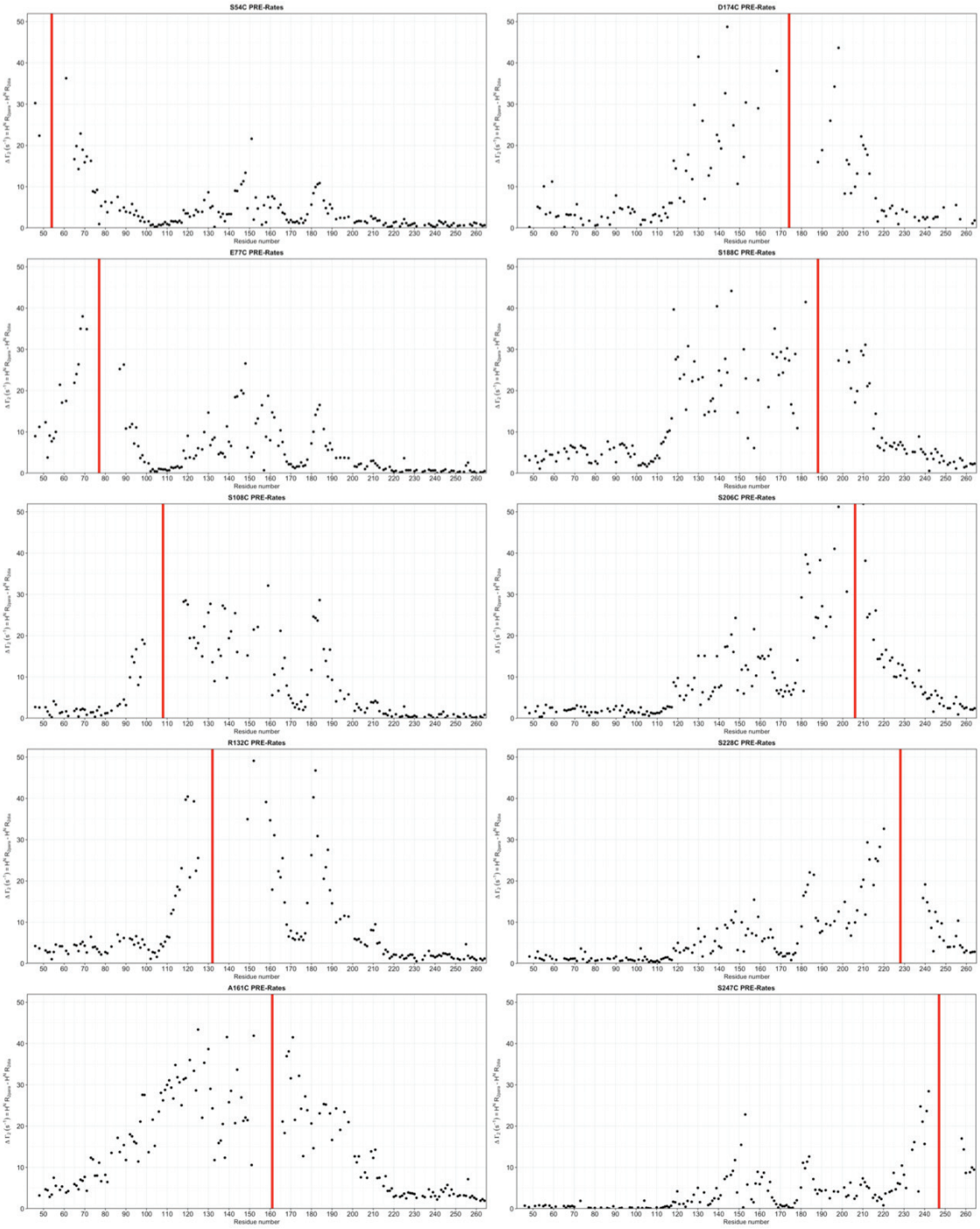

Figure S3. All PRE-Profiles of qOPN - Red bars indicate the position of the spin label.

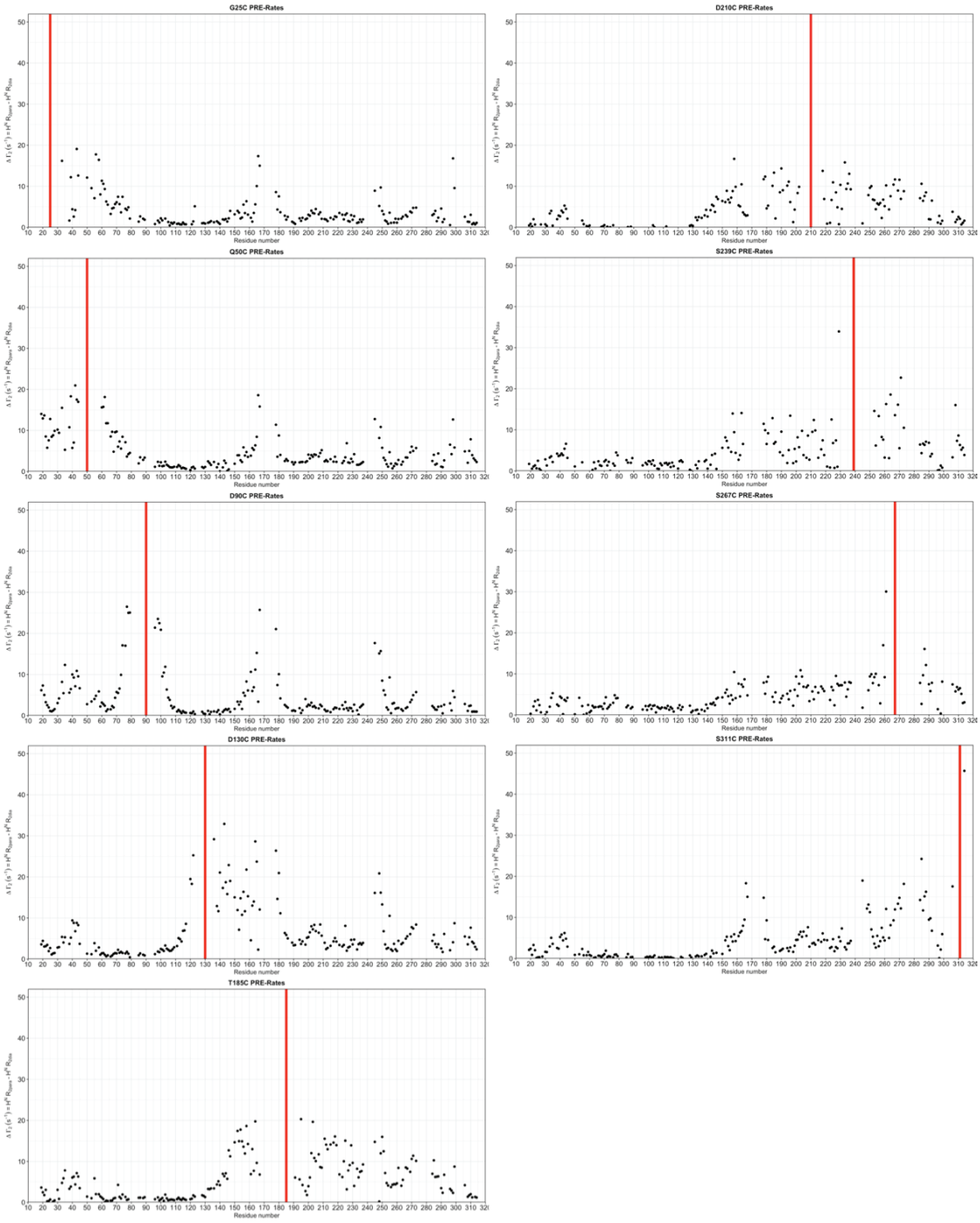

Figure S4. All PRE-Profiles of hOPN - Red bars indicate the position of the spin label.

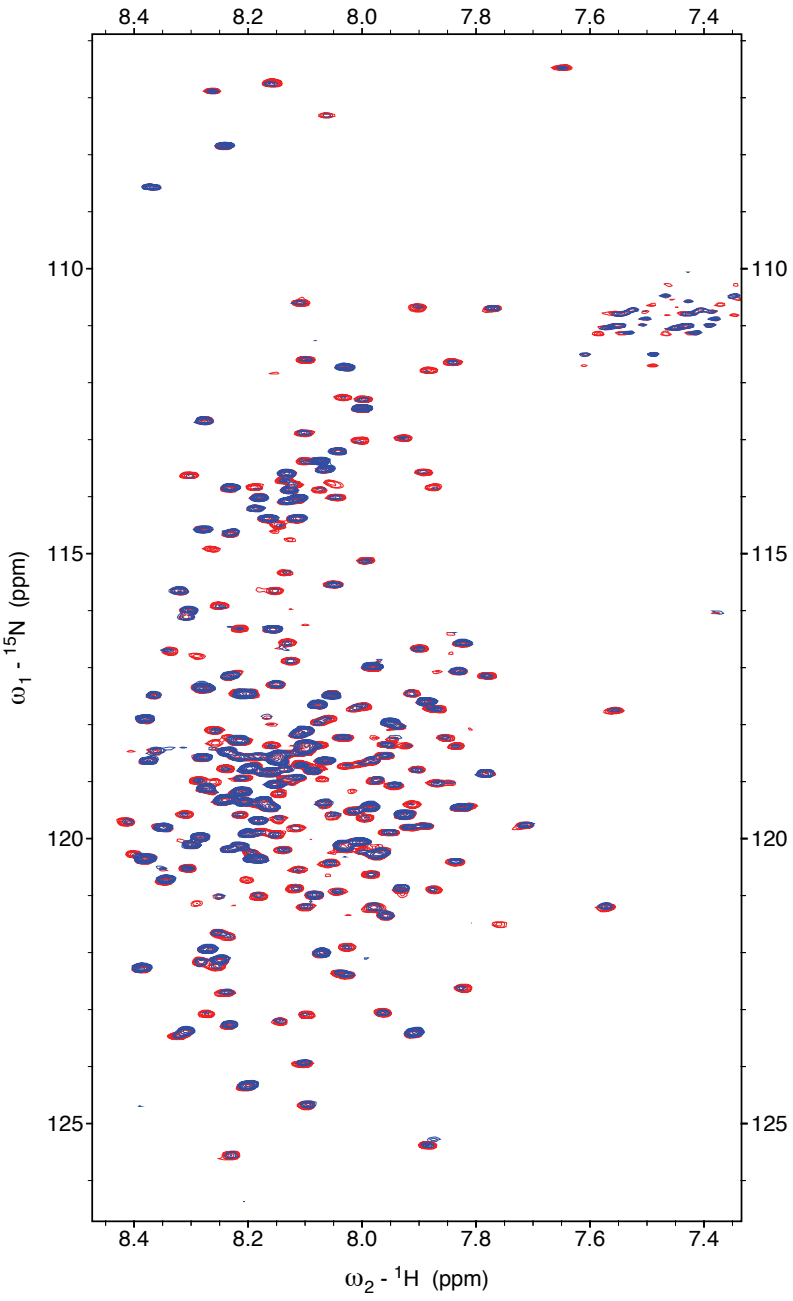

Figure S5. The NMR Spectra of hOPN before (red) and after (blue) binding to CD44.

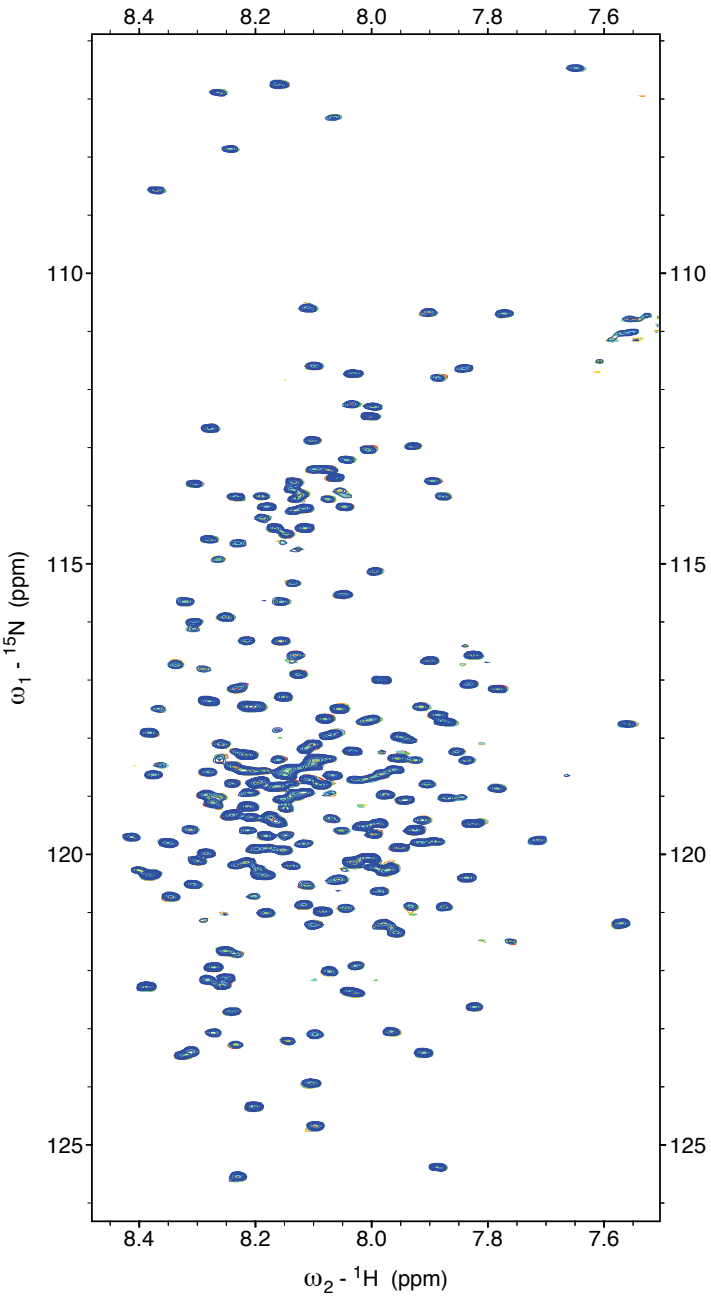

Figure S6. Titration of hyaluronic acid with hOPN.

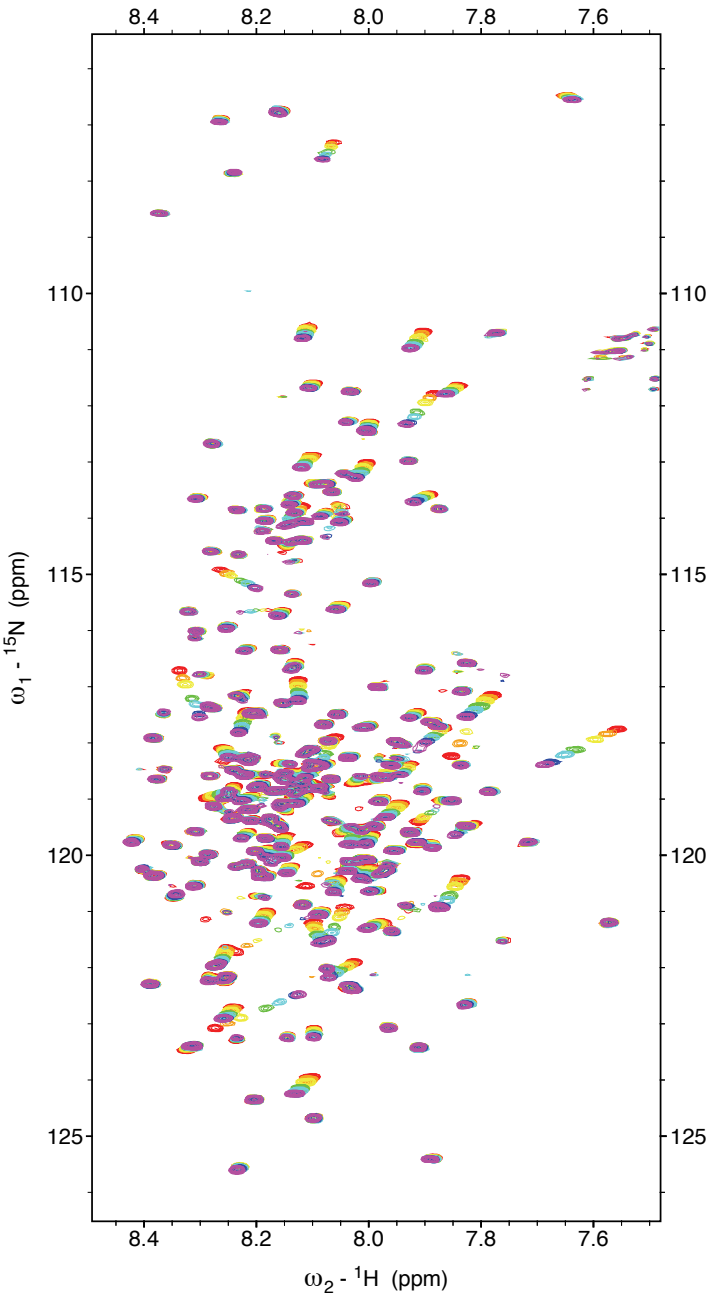

Figure S7. Titration of heparin with hOPN.

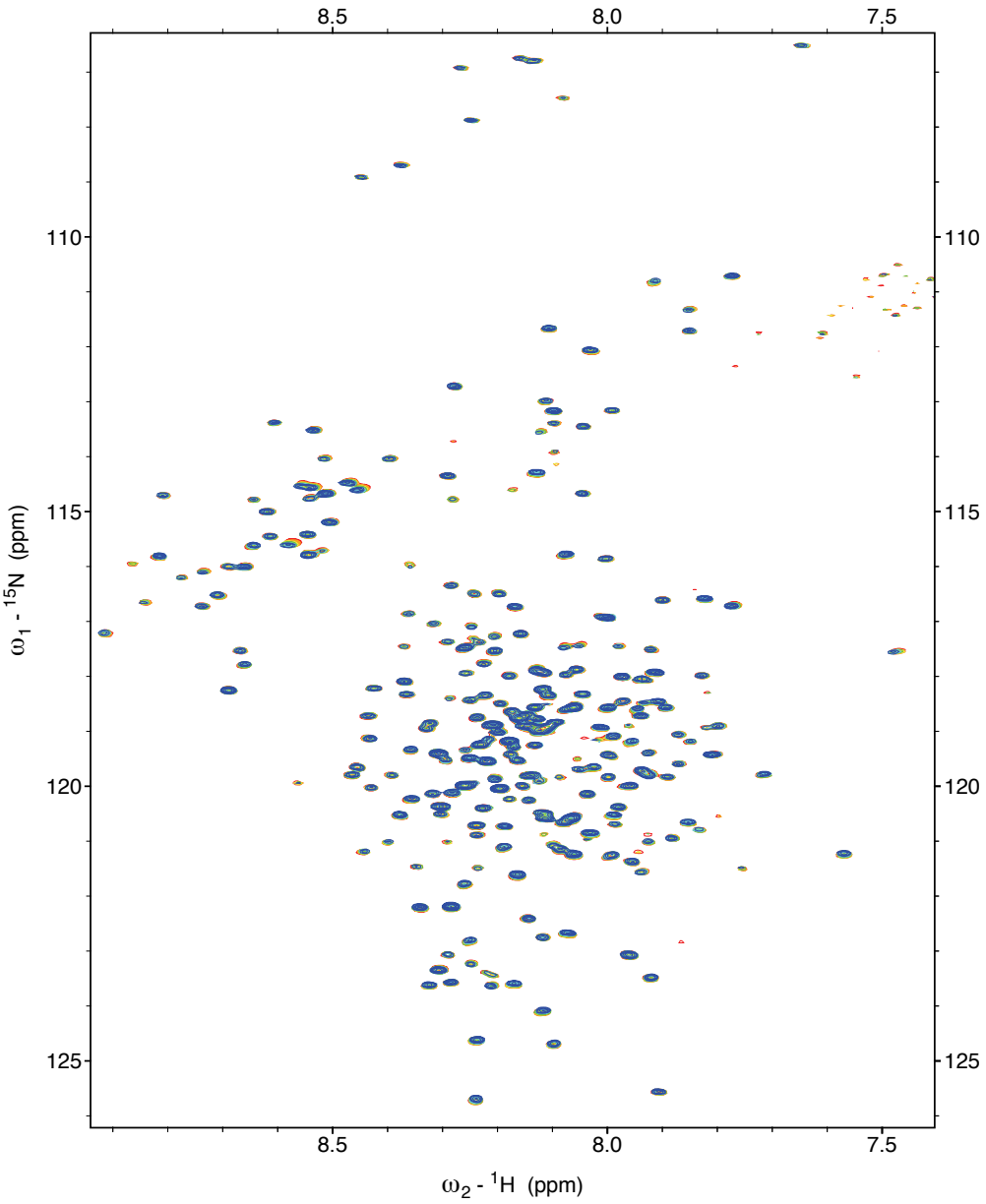

Figure S8. Titration of hyaluronic acid with phosphorylated hOPN.

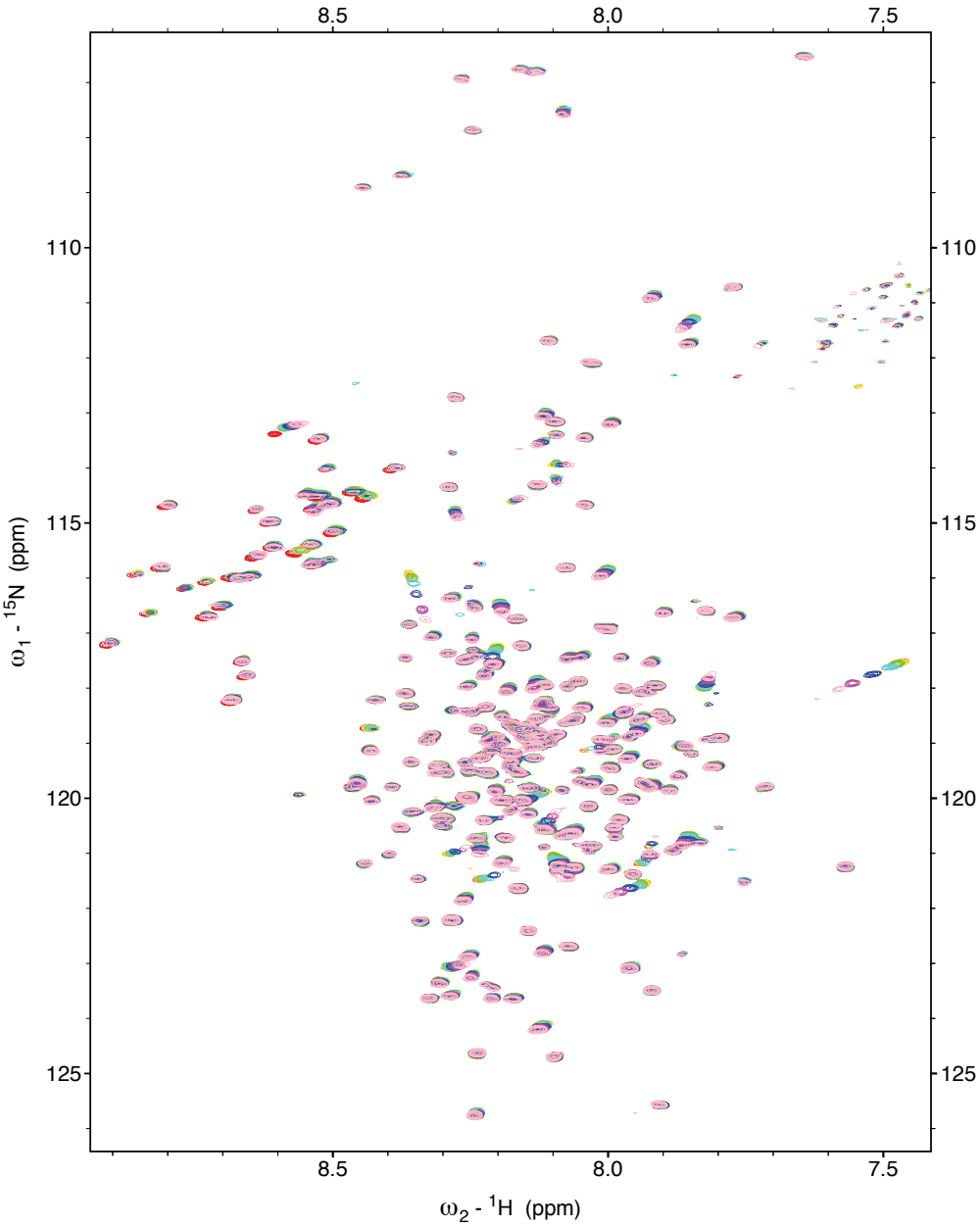

Figure S9. Titration of heparin with phosphorylated hOPN.

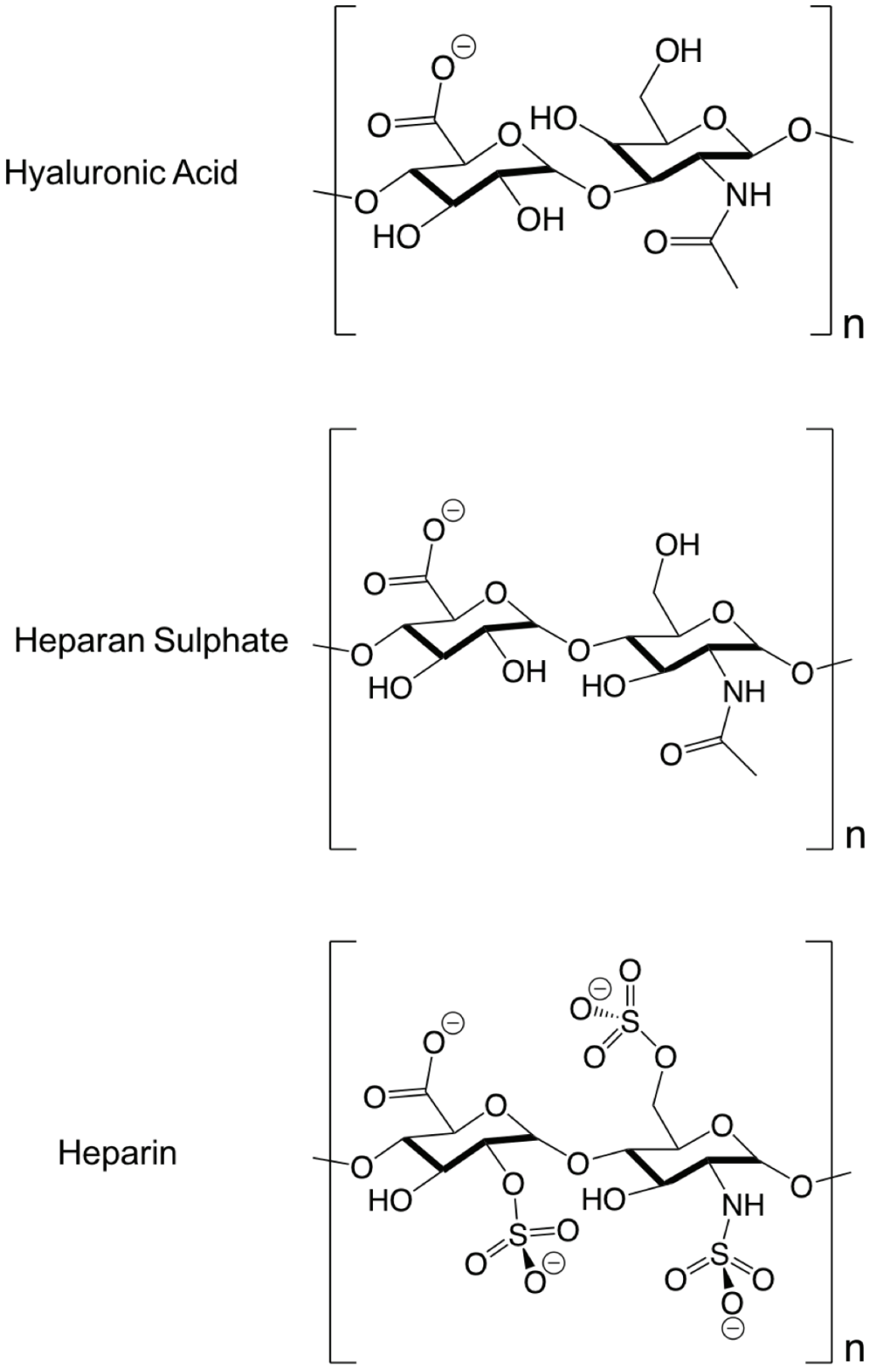

Figure S10. The basic building blocks of Hyaluronic Acid, Heparan Sulphate and Heparin. Heparan sulphate is variably sulfated, adding negative charges to the biomolecule.

Abbreviations

ATP	adenosine triphosphate
CSP	chemical shift perturbation
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
Fam20c	family with sequence similarity 20 member C
HABD	hyaluronic acid binding domain
HEK cells	human embryonic kidney cells
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
IPTG	Isopropyl-β-D-Thiogalactopyranoside
LB	lysogeny broth medium
MES	2-(N-morpholino)ethanesulfonic acid
MSA	multiple sequence alignment
MTSL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl) methyl methanesulfonothioate
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600nm
OPN	osteopontin
PBS	phosphate-buffered saline
PRE	paramagnetic relaxation enhancement
PTM	post-translational modification
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TEV protease	tobacco etch virus protease
Tris	tris(hydroxymethyl)aminomethane