universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

The intrinsically „not-so-unfolded“ protein Osteopontin

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer.nat.)

Verfasserin / Verfasser:	Gerald Platzer
Matrikel-Nummer:	0305853
Studienrichtung (lt. Studienblatt):	Molekulare Biologie
Betreuerin / Betreuer:	Robert Konrat

Wien, am 30.11.2009

1 Table of contents

2 Abstract

Osteopontin (OPN) is a member of the small integrin binding ligand N-linked glycoprotein (SIBLING) family of proteins. The molecular structure of OPN contains unique structural motifs, which mediate cell-matrix and cell–cell signaling through the αvβ integrin family and CD44 receptors in a variety of normal and pathologic processes.

Avian fibroblasts simultaneously transformed by *v-myc* and *v-mil(raf)* display significantly elevated levels of the transcription factor complex AP-1 which subsequently leads to upregulation of the osteopontin (OPN) encoding gene (*OPN*). Elevated OPN expression is commonly observed in tumor cells where it plays crucial roles in remodeling processes such as inflammation, bone resorption and tumor progression.

OPN belongs to the class of intrinsically unfolded proteins (IUP) with no apparent ordered tertiary structure detectable. Intrinsically unfolded proteins have challenged the structure–function paradigm in current biology, since they are now recognized to play essential roles in transcription, translation and signal transduction pathways. The absence of defined structure allows disordered regions of proteins to acquire a high degree of plasticity but also represents an easy target for misfolding, often leading to disease.

NMR is uniquely suited to elucidate the structural and dynamic features of these unfolded states. A combination of ^{15}N-relaxation time measurements, pulsed-field gradient (PFG) diffusion measurements, residual dipolar couplings and paramagnetic relaxation enhancement experiments were applied to characterize the dynamic properties of OPN. This work will address the connection between the molecular structure of Osteopontin and its versatile physiological functions.

3 Zusammenfassung

Osteopontin (OPN) ist ein Mitglied aus der Familie der small integrin binding ligand N-linked glycoproteins (SIBLINGs). Seine molekulare Struktur beinhaltet Motive, die für Zell-Zell und Zell-Matrix Signaling via der αvβ-Integrin- oder CD44 Rezeptoren in normalen sowie pathologischen Prozessen notwendig sind.

v-myc and *v-mil(raf)* co-transformierte Hühner Fibroblasten besitzen eine erhöhte Konzentration des Transkriptionsfaktors AP-1, welcher zu einer erhöhten Expression des für OPN kodierenden Gens *OPN* führt. Eine erhöhte OPN Expression ist häufig in Tumorzellen zu beobachten, wo das Protein an einer Reihe von Prozessen wie Remodeling, Inflammation, Knochen Resorption und Tumorwachstum.

OPN ist ein sogenanntes intrinsically unfolded protein (IUP) mit keiner definierten Tertiärstruktur. IUPs stellen das gängige Struktur-Funktions Paradigma der Biologie in Frage. Trotz des Fehlens einer definierten Struktur sind sie an wichtigen Prozessen wie Transkriptions, Translations und Signaltransduktionsprozessen beteiligt. Seine konformationelle Flexibilität macht es jedoch auch zum Ziel von Misfolding was häufig zu Krankheiten führt.

NMR eignet sich hervorragend dazu, die Dynamik und Struktur solcher ungefaltenen Proteine genauer unter die Lupe zu nehmen. Eine Kombination aus ^{15}N-Relaxationsmessungen, pulsed-field gradient (PFG) diffusion measurements, residual dipolar couplings (RDCs) und paramagnetic relaxation enhancement (PRE) Experimenten wurde verwendet, um den dynamischen Charakter von OPN genauer zu untersuchen. Diese Arbeit versucht, die molekulare Struktur von OPN mit seinen vielschichtigen physiologischen Aufgaben in Verbindung zu bringen.

4 Introduction

4.1 The SIBLING family

Small integrin-binding ligand N-linked glycoproteins (SIBLINGs) are a family of five integrin-binding glycophosphoproteins comprising osteopontin (OPN), bone sialoprotein (BSP), dentin matrix protein 1 (DMP1), dentin sialophosphoprotein (DSPP) and matrix extracellular phosphoglycoprotein (MEPE).

Each of the SIBLING genes map to human chromosome 4 and all of the SIBLING proteins contain the integrin binding motif, Arg–Gly–Asp (RGD) [1]. SIBLING expression was first evidenced in mineralized tissue including bone and teeth. Recent studies have demonstrated that SIBLINGs localize to neoplastic tissues and function in mediating metastasis.

All of the SIBLING proteins are extended and flexible in solution, a property shared by numerous proteins with multiple binding partners. Consistent with that observation, SIBLINGs bind to a number of different protein families, including integrins (through both RGD and non-RGD motifs), other cell-surface proteins (e.g. CD44), complement factor H (CFH) and members of the matrix metalloproteinase (MMP) family.

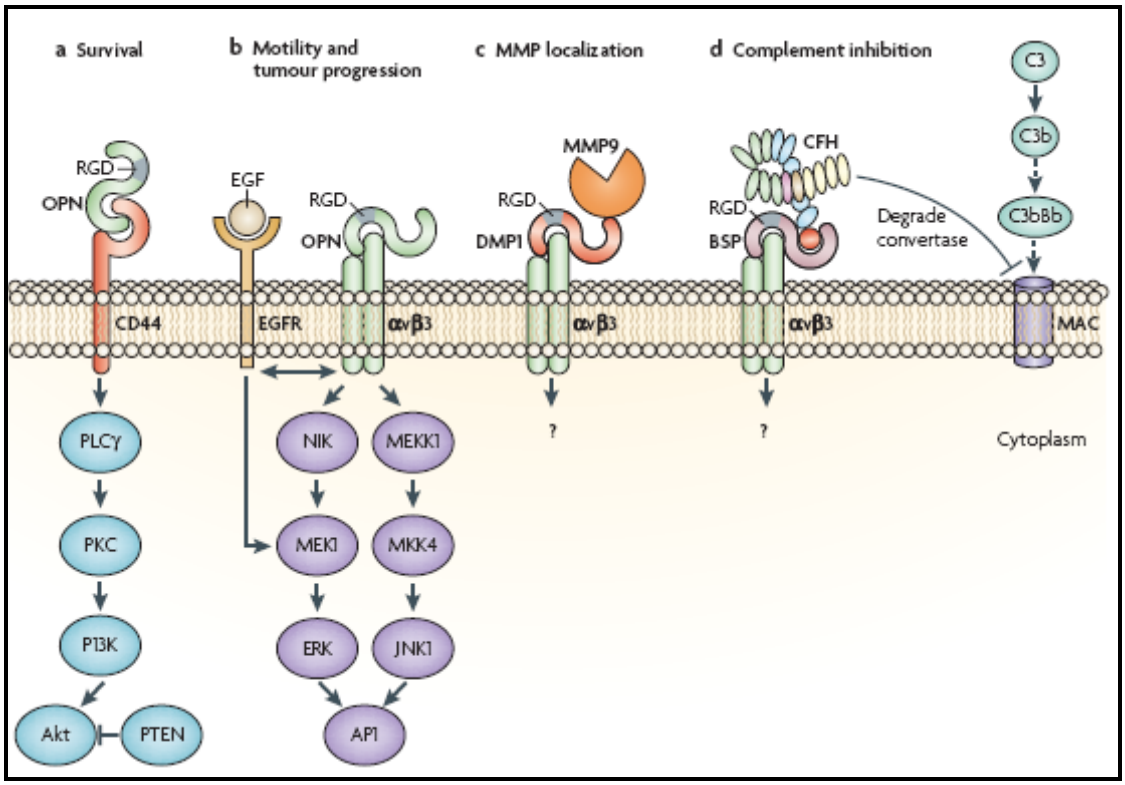


Figure 1: SIBLINGs mediate cell–matrix interactions and cellular signalling.

These interactions enable cell surface localization and release of MMPs and CFH by at least OPN, BSP and DMP1, thereby enabling their biological functions (e.g.: extracellular matrix degradation and evasion of complement-mediated lysis). Some members of the SIBLING family were discovered as abundant proteins trapped within the mineralized tissue of bone and dentin [2],[3]. In the early years of research, each of these proteins was thought to be both skeletal tissue-specific and to have a role in nucleating hydroxyapatite crystals within mesenchymal tissues. [4] From the 1990s, various combinations of SIBLING proteins were discovered to be significantly upregulated in a number of epithelial tumors known to metastasize to bone [5]. SIBLINGs are secreted proteins that are localized through interactions with receptors either on their own cell surface, enabling autocrine activities, or by diffusing short distances to surrounding cells where they may function in paracrine signaling. SIBLINGs propagate biological signals by initiating integrin signaling and by recruiting other proteins to the cell surface.

4.2 SIBLINGs and tumor progression

Tumor progression involves a series of events that grants transformed cells a higher survivability. These events start with neoplastic transformation and continue with the deactivation of proliferation blockades, growth restriction, physical barriers and host defence systems. Cancer cell survival is dependent on proliferation, interaction with the extracellular matrix to create space to grow, pathways for nutrient access and escape of the cells to a new environment. Successful progression also involves cellular responses and evasion of immune surveillance.

4.2.1 Survival and proliferation

At the primary site, cancer cells secrete high levels of small integrin-binding ligand, N-linked glycoproteins (SIBLINGs), which favor their proliferation (osteopontin (OPN) and bone sialoprotein (BSP)) and survival (OPN, BSP and dentin matrix protein 1 (DMP1)).

Cancer cells bind to SIBLINGs and their various proteolytic fragments through their integrin receptors by both RGD-dependent and RGD-independent interactions. OPN, and perhaps DMP1, can also interact with specific splice variants of CD44 that are expressed by cancer cells.

Osteopontin is the best-studied protein among the SIBLING family and the intracellular signaling pathways present in OPN modulation of cell proliferation and migration are well understood. The binding of OPN to CD44 promotes cell migration through kinase cascades involving phospholipase Cγ, protein kinase C, phosphatidylinositiol 3-kinase (PI3K) and Akt, a serine/threonine kinase that regulates cell cycle progression, growth factor-mediated survival and cell migration (Figure 2a). The binding of αvβ3 by OPN is associated with SRC kinase-mediated complex formation between αvβ3 and EGFR, which activates the mitogen-activated protein kinase (MAPK) pathway and results in the promotion of tumor growth (Figure 2b). The effects of SIBLINGs other than BSP and OPN on cell proliferation have not been investigated yet.

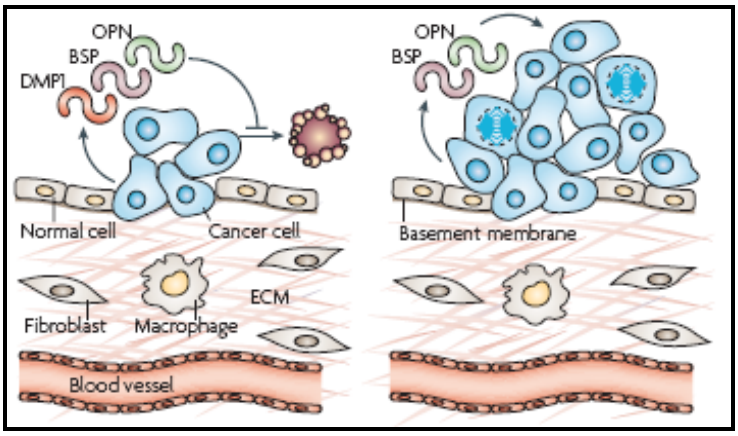


Figure 2: Survival / Resistance to Apoptosis and Proliferation

4.2.2 Invasion and ECM degradation

Cancer cells exhibit a high motility in combination with increased expression of proteases that are able to degrade the ECM thus providing them with a high invasive capability [6],[7]. Both OPN and BSP are expressed at increased levels by several cancers contributing to their invasive potential.

Invasive cells can degrade the extracellular matrix through at least two pathways of controlled proteolysis: the urokinase-type plasminogen activator (uPA) pathway and the MMP (matrix metalloproteainase) pathway (Figure 3).

Several studies with recombinant OPN show that this SIBLING member significantly increases *in vitro* invasiveness:

In a metastatic murine mammary cancer cell lines model, the binding of OPN to integrin receptors induces MMP2 and uPA expression through integrin linked kinase (ILK)-dependent AP1 activity [8]. It has recently been shown that OPN induces αvβ3 integrin-mediated AP1 activity and uPA secretion by activating SRC–EGFR–ERK (extracellular signal-regulated kinase) signaling pathways and further demonstrates a functional molecular link between OPN-induced integrin- and SRC-dependent EGFR phosphorylation and ERK- and AP1-mediated uPA secretion, and all of these ultimately control the motility of breast cancer cells [9]. Potential mechanisms for SIBLING-enhanced matrix degradation have been described. BSP and OPN induce the activation of MMP2 in GCT23 giant cell tumour cells [10]. OPN binding to αvβ3 is associated with

PI3Kmediated NFκB activation and nuclear factor-inducing kinase (NIK, also known as mitogen-activated protein kinase kinase kinase 14 (MAP3K14)) activation of AP1 and NFκB, which stimulate uPA-dependent MMP9 activation [11]. Thus, at least two SIBLINGs can induce MMP expression. BSP, DMP1 and OPN bind to and modulate the activity of MMP2, MMP9 and MMP3, respectively [12]. SIBLING-mediated MMP activation includes both making the proMMPs enzymatically active to some degree and reactivating MMPs that are inhibited by tissue inhibitors of MMP (TIMPs) [12]. Another mechanism might involve enzymatic processing of SIBLINGs that alters invasion and migration properties. For example, increased hepatocellular carcinoma cell invasion was attributed to OPN peptides cleaved by MMP9 and thrombin [13].

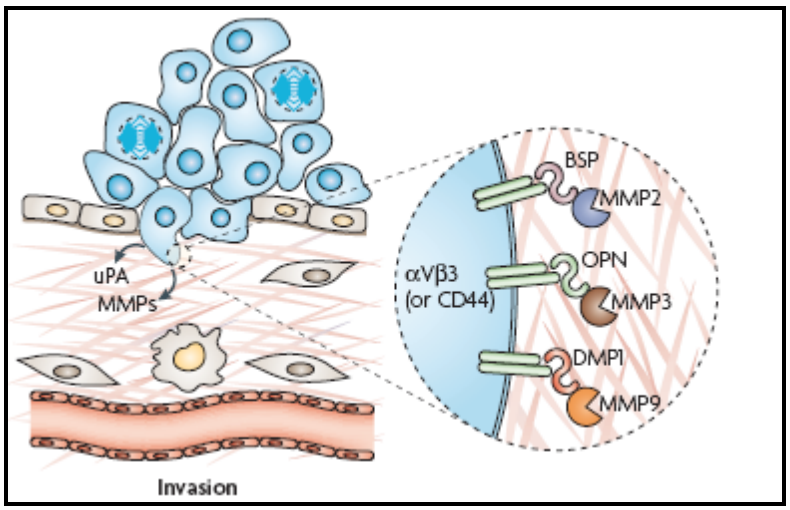


Figure 3: Tumor invasion of surrounding tissue

4.2.3 Metastasis

The process of metastasis is very complex and is split into several stages: The proliferation of malignant cells at the primary site is followed by spreading, to invade surrounding capillary or lymphatic vessels. Tumor cells are able to resist immunological attacks and eventually gain access to secondary sites where they proliferate to form a new tumor [14]. Throughout this multi-step progression, cancer cells interact with ECM-proteins and various types of cells. Multifunctional ECM-proteins such as the SIBLINGs are anticipated to have key roles in metastasis as they affect adhesion, migration and matrix degradation (Figure 1).

OPN was the first member of the SIBLING family found to confer a metastatic phenotype on cells that originally formed benign tumors [15]. Since this initial observation, several studies conducted with various types of cancer from human tissue report that tumors that are likely to progress to more advanced stages exhibit *de novo* or increased expression of SIBLINGs. In particular, the correlations between high levels of OPN expression in tumor cells and its pro-metastatic role have been observed. So far, high expression of SIBLINGs is associated with osteotropic cancers including breast [16], prostate [17] and lung [18] as well as multiple myeloma [19].

4.2.4 Inflammation and complement evasion

A very interesting feature of SIBLING proteins is their function as part of the immune system. OPN is secreted by activated macrophages, leukocytes and activated T lymphocytes [20-22] and is also a chemotactic cytokine for macrophages [23], dendritic cells [24] and neutrophils [25]. It is possible that OPN expression by tumors actually promotes inflammation induced cancer growth and progression through, for example, the promotion of macrophage and neutrophil infiltration. Tumor cells that secrete OPN might be propagating chronic inflammation, which can accelerate transformation and tumor progression.

The complement system is yet another part of the immune system SIBLING proteins are involved in. The complement system is composed of about 26 proteins that combine with antibodies and/or cell surface molecules as part of the humoral response. The complement cascade has a role in inflammation, opsonization, viral neutralization, cell lysis and localization of antigen [26]. Nearly all cells are subject to constant low levels of complement attack, but only cells that do not express the correct cell surface proteins, thereby inactivating the early steps of the cascade, are killed. CFH is a major negative regulatory factor that quenches complement- mediated lysis. Cells that become transformed may escape the complement system during their transit through the patient's circulation by upregulating genes that help to control this aspect of immune surveillance. As such, the expression of SIBLINGs (specifically OPN, DMP1 and BSP) by tumor cells might provide such a selective advantage for survival by mediating the binding of CFH to the cell surface through integrins and/or CD44. The activated CFH then inhibits the

formation of the membrane attack complex and subsequent cell lysis (Figure 4). *In vitro* experiments have demonstrated that these three SIBLINGs can protect murine and human cancer cell lines from attack by complement [27-29].

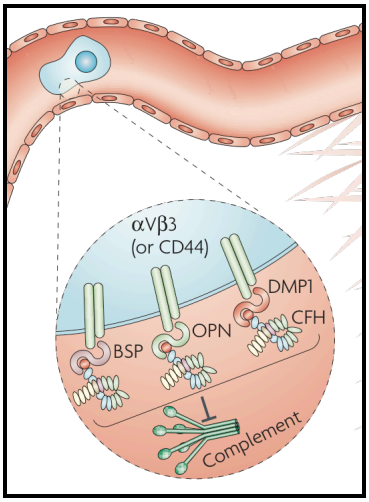


Figure 4: Complement evasion

4.2.5 Angiogenesis

Angiogenesis promotes tumor growth as well as metastatic spread through a complex interplay of positive and negative mediators of extracellular matrix degradation and endothelial cell and vasculature recruitment. There is evidence that αvβ3 is a key angiogenesis-associated receptor that is significantly upregulated on the surface of activated endothelial cells [30]. OPN and BSP have been shown to act as pro-angiogenic factors and, based on their RGD motifs, it is likely that the other SIBLINGs may also interact with αvβ3 integrin and influence the behavior of endothelial cells. OPN contributes to the genesis of new capillaries infiltrating the cancer lesion in several *in vivo* models [31],[32]. The integrin αvβ3 mediates the migration of activated endothelial cells during vessel formation. SIBLINGs, as ligands for αvβ3 through the RGD sequence, may stimulate endothelial cell migration. It is also possible that SIBLING modulation of protease activity (uPA or MMP) generates bioactive fragments of extracellular matrix components responsible for angiogenesis. Experimental evidence suggests that antagonizing the ligation of SIBLINGs to integrins is a promising approach for the inhibition of angiogenesis and associated tumor growth. For example, blocking the

interaction between OPN and αvβ3 inhibits angiogenesis and stops lung cancer growth in mice [31]. The αvβ3 integrin was shown to be important for OPN-mediated NFκB induction and survival, as adding a neutralizing anti-β3 integrin antibody blocked NFκB activity and induced endothelial cell death when cells were plated on OPN [33]. A recent study demonstrates that OPN triggers vascular endothelial growth factor-dependent breast tumor growth and angiogenesis by autocrine and paracrine mechanisms [34]. Thus, it is possible that, through their interaction with αvβ3, the SIBLINGs may also cooperate with molecules that have important biological functions during angiogenesis and tumoral growth processes, including MMPs, growth factors and their receptors.

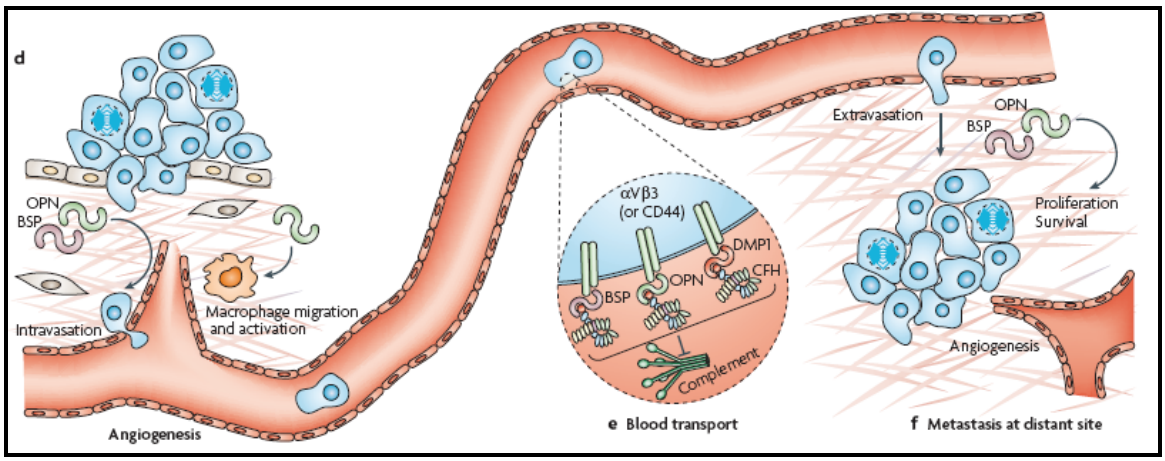


Figure 5: Angiogeesis, Blood Transport and Metastasis

4.2.6 Microcalcification

All SIBLINGs are expressed by bones and teeth. Outside of the skeletal system, pathological dystrophic calcification associated with the upregulation of OPN and/or BSP has been observed. Notable among these are atherosclerotic vascular plaques and kidney stones [35]. Because of their earlier association with mineralization in bone, the expression of BSP and OPN has been studied in cancers such as breast and thyroid carcinomas, in which ectopic calcification occurs [36],[37]. Tumors from the primary sites of bone-seeking cancers frequently contain foci in the form of hydroxyapatite (HA) microcalcifications.

OPN is capable of both inhibiting and stabilizing the formation and growth of hydroxyapatite and other biominerals. The interaction of OPN with HA is determined by the extent of protein phosphorylation and that this interaction regulates the mineralization process. [38] For example recombinant, nonphosphorylated OPN and chemically dephosphorylated OPN, has no effect on HA formation or growth. In contrast, highly phosphorylated milk OPN promoted HA formation. A feature of many of the anionic extracellular matrix proteins that regulate initial formation and growth of HA crystals and other biominerals [39][40] is their ability at low concentrations to stabilize the first formed "critical" HA nuclei facilitating initial mineral and, at higher concentrations, to bind to crystals limiting their growth and proliferation.

4.3 Molecular structure of OPN

In this work, a truncated 220 amino acid long OPN version from quail was used for all experiments. From now on it will be referred to as OPN200 in all the subsequent data.

OPN exhibits a high degree of sequence homology among species. (Figure 6: Multiple Sequence Alignment of OPN220 with human-OPN and native quail-OPN. Conserved regions are highlighted in blue. Important binding motifs are boxed in red.)

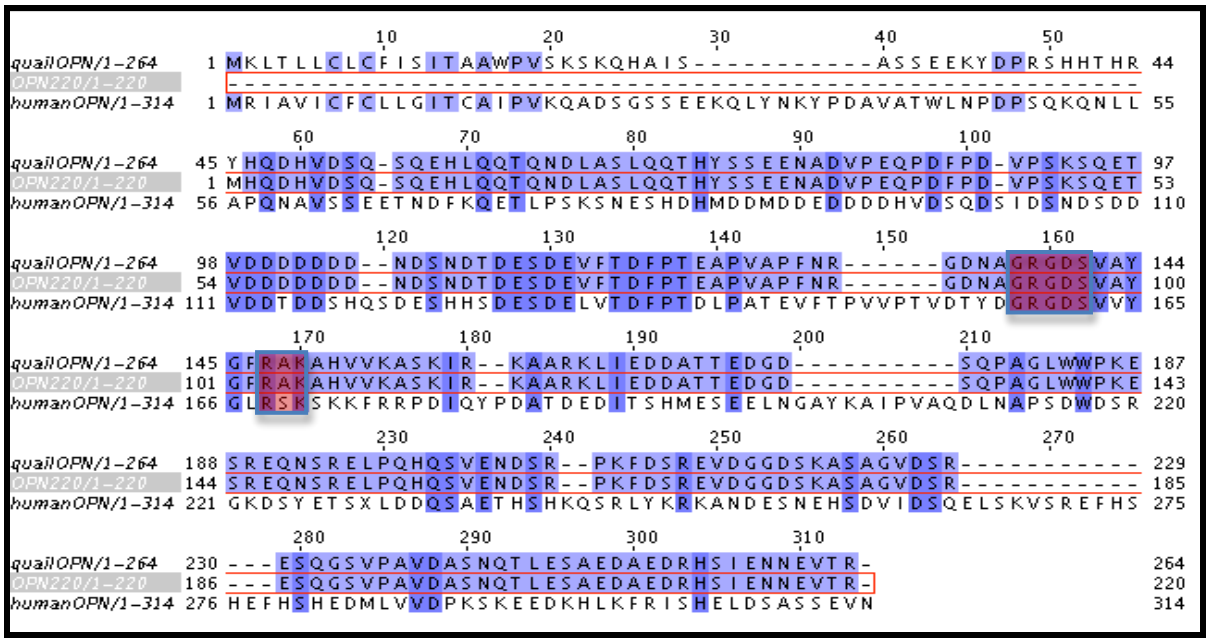


Figure 6: Multiple Sequence Alignment of OPN220 with human-OPN and native quail-OPN. Conserved regions are highlighted in blue. Important binding motifs are boxed in red.

OPN is rich in aspartate, glutamate and serine residues and contains functional domains for extra-cellular matrix adhesion. (Figure 7: Schematic representation of the domain structure of OPN220.Figure 7) It interacts with a variety of integrins, including $\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_v\beta_1$, $\alpha_4\beta_1$, $\alpha_8\beta_1$ and $\alpha_9\beta_1$, as well as CD44.

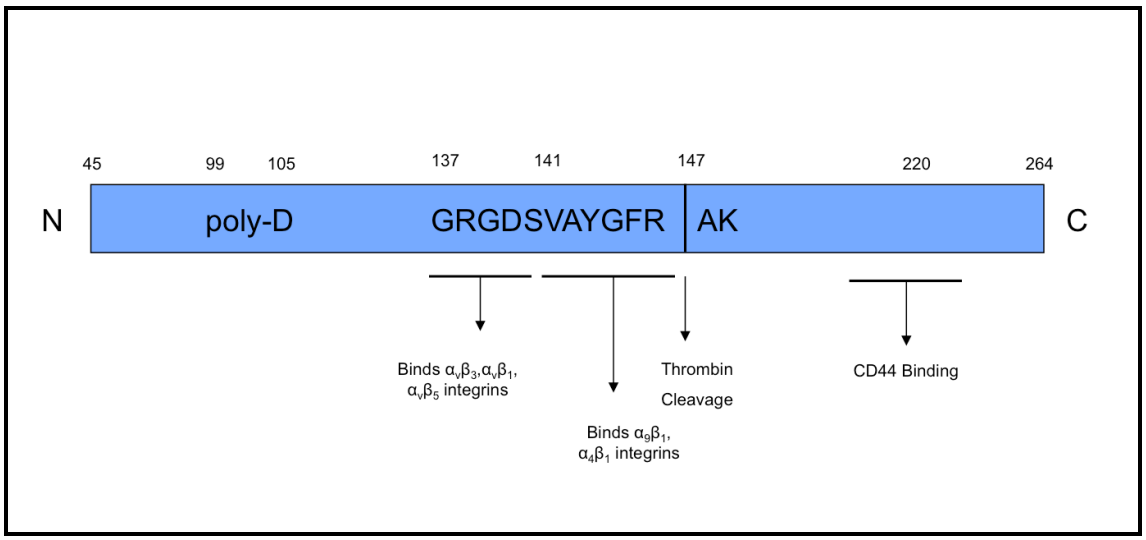


Figure 7: Schematic representation of the domain structure of OPN220.

Integrin mediated cell adhesion and migration are stimulated in assays in which full-length OPN is immobilized on tissue culture dishes. Thrombin cleavage of the protease hypersensitive RSK (RAK in quail) sequence separates the integrin- and CD44-binding domains, which in some cases promotes adhesion over migration. For example, the thrombin-generated amino-terminal OPN fragment binds to αvβ3 and αvβ5 integrins (through the RGD motif [41]) or to α9β1 and α4β1 integrins (through the cryptic SVVYGLR sequence which is only partly conserved in quail [42]) and promotes cell adhesion. The carboxy-terminal fragment binds to CD44 and promotes invasion, tumorigenesis and the formation of foci [43].

OPN contains a polyaspartic acid sequence (D99-D105 in OPN220) and sites of Ser/Thr phosphorylation, of which half are highly conserved. In non-recombinant proteins, these residues are postulated to mediate OPN binding to hydroxyapatite [44]. Human OPN also contains transglutaminase-reactive glutamine residues (Q34 and Q36) not found in quail-

OPN that allow it to be cross-linked by transglutaminase enzymes. This cross-linking results in the formation of OPN polymers, which also bind to HA [45][46][47].

4.4 The advantage of being intrinsically unstructured

In the recent years, a large number of papers on proteins denoted as 'natively unfolded' or 'intrinsically unstructured/disordered' have appeared. The term intrinsically unfolded can be misleading, because rather than having no structure, these proteins sample a lot of conformations with no distinct energy minimum, which is in stark contrast to globular proteins. Unstructured proteins have no enzymatic activity, as a proper organization of the active site residues requires a well-structured fold they cannot provide, but exhibit order at the level of primary and secondary structure that correlates with their functions. The lack of a stable fold and the exhibition of a rather extended conformation provide these proteins with a high intramolecular flexibility.

IUPs differ significantly from average genomically encoded proteins in terms of size and amino acid composition. (Figure 8: Amino Acid Composition of OPN220)

	number of residues within OPN	frequencies of residues within OPN
Ala (A)	19	8.64
Arg (R)	12	5.45
Asn (N)	11	5.00
Asp (D)	32	14.55
Cys (C)	0	0.00
Gln (Q)	16	7.27
Glu (E)	21	9.55
Gly (G)	10	4.55
His (H)	7	3.18
Ile (I)	3	1.36
Leu (L)	7	3.18
Lys (K)	9	4.09
Met (M)	1	0.45
Phe (F)	6	2.73
Pro (P)	12	5.45
Ser (S)	25	11.36
Thr (T)	10	4.55
Trp (W)	2	0.91
Tyr (Y)	2	0.91
Val (V)	15	6.82

Amino acid frequencies for quail OPN220.

The high abundance of proline and charged amino acids as well as the lack of cysteins is a shared characteristic among IUPs.

Total number of negatively charged residues (Asp + Glu): 53
Total number of positively charged residues (Arg + Lys): 21

Figure 8: Amino Acid Composition of OPN220

A low mean hydrophobicity and high net charge which are associated with the disordered state [48][49], preclude the formation of a hydrophobic cluster and promote an extended conformation by electrostatic repulsion. The lack of cysteine fits into this picture, as in globular proteins this amino acid is often found in active sites or stabilizing disulfide bonds that are not required in IUPs. High proline content is also linked with the lack of structure. Proline disfavors a stable tertiary structure but has a strong preference for an extended conformation in the form of polyproline II helices. Furthermore, proline is often involved in protein–protein contacts [50], supporting the high abundance of this amino acid in IUPs, which heavily rely on target recognition.

Another unique functional feature of IUPs is that their open structure is largely preserved when they complex with their target, resulting in a large binding surface and multiple contact points for a protein of the given size. For effectors, it provides specificity because of interaction with distant regions on the target, as can be seen with the X-ray structure of the cyclin-dependent-kinase inhibitor p27Kip1 bound to the cyclin A–cyclin dependent kinase 2 (Cdk2) complex (Figure 9).

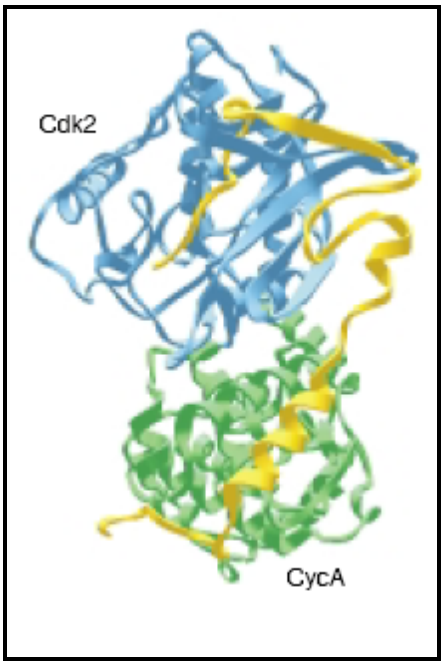


Figure 9: Cyclin-dependent-kinase inhibitor p27Kip1 bound to the cyclin A–cyclin dependent kinase 2 (Cdk2) complex

The high plasticity of IUP's allows them to adopt different structures upon different stimuli or with different partners, which enables their versatile interaction with various targets. Such is exactly the case for the SIBLING proteins, which can interact with integrins, CD44 and complement factor H. The lack of a stable tertiary fold and the high plasticity goes in hand with a high degree of intramolecular dynamics compared to stably folded proteins, a feature of OPN which was of particular interest to me.

NMR is uniquely suited to elucidate these structural and dynamic features. A combination of NMR experiments including ^{15}N-relaxation time measurements, pulsed-field gradient (PFG) diffusion measurements, residual dipolar couplings and paramagnetic relaxation enhancement experiments in combination with SAXS and CD were applied to characterize global and dynamic properties of OPN.

5 Results

Osteopontin is a very interesting protein by all means. Structure determination by conventional NMR experiments is not possible, due to the unstructured nature of all of the SIBLING members. Nevertheless it is possible to get some insights on the dynamic properties of intrinsically unstructured proteins.

During the past decade numerous NMR techniques have been developed for elucidating the unfolded states of proteins in atomic detail [51]. 15N-relaxation time measurements and paramagnetic relaxation enhancement (PRE) from site-directed spin labeling allowed the detection of native-like contacts and hydrophobic clusters in proteins [52][53]. In addition, measurement of residual dipolar couplings (RDCs) in a weakly aligned protein provides long-range orientational information and may reveal residual native-like topology in the unfolded states of proteins [54][55].

The first half of the results part will deal with a detailed analysis of global (size and composition of secondary structure elements) as well as dynamic properties (T1, T2, heteronucular NOE, RDC's, PRE's) of osteopontin and the corresponding NMR experiments this data was extracted with.

The second part will be concentrated on another interesting feature of IUPs. They are almost completely devoid of a folded structure, therefore exhibiting a very high binding plasticity. The fact that Osteopontin is known to bind a number of ligands including integrins, CD44 and MMPs made us curious. Hence we decided to make use of Robert Konrats *Meta-Structure* Algorithm to search for putative ligands for osteopontin.

The *Meta-Structure* concept was developed by Robert Konrat with the aim to extract information about protein structure and function out of the primary sequence of a protein without the help of 3D structural information, which often, even at high resolutions, is insufficient to predict protein function or functional sites. Individual residues are treated as tightly interacting chemical entities and their mutual interdependences can be quantified using parameters borrowed from network theory. Meta-structure analysis from the primary sequence of a protein gives information about the compactness of a protein (whether proteins are highly structured, unstructured or partially folded). It can be used for the search of protein fold topology similarities (meta-structure homologues), which

are not necessarily visible in the primary sequence of proteins. Furthermore, it is able to predict protein interaction and ligand binding sites [56].

5.1 Secondary Chemical Shifts

The differences between observed chemical shifts of nuclei spin types and their random coil values assigned in unfolded conformations serve as feasible tool for evaluating the residual preferences for protein secondary structure. Generally, backbone $^{1}H^{N}$, ^{15}N, $^{13}C\alpha$, and side-chain $^{13}C\beta$ chemical shifts were used, using a set of established corrections factors.

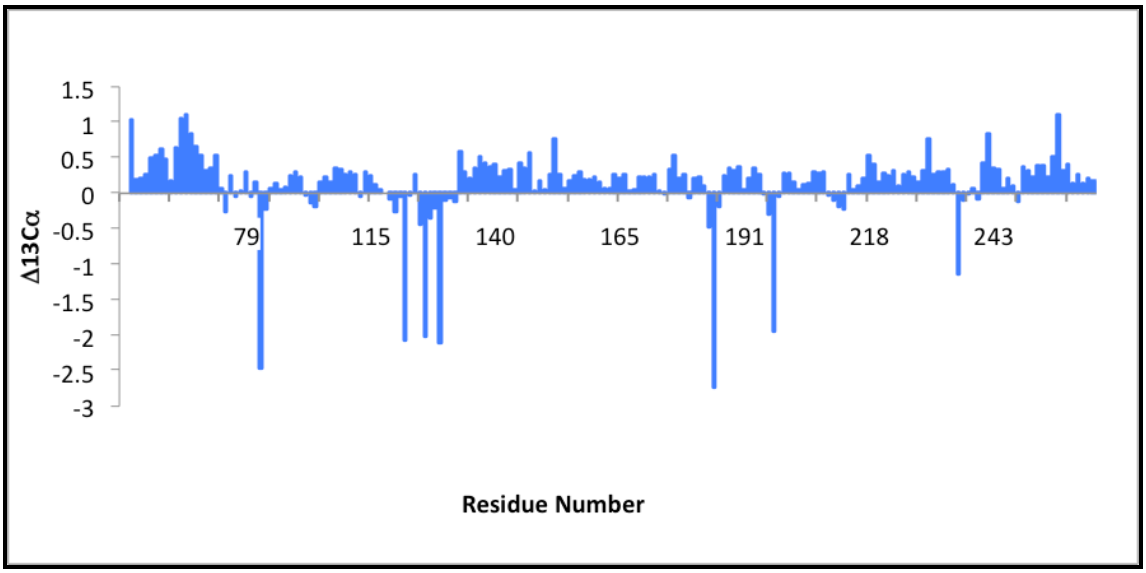


Figure 10: OPN220 $^{13}C\alpha$ chemical shifts

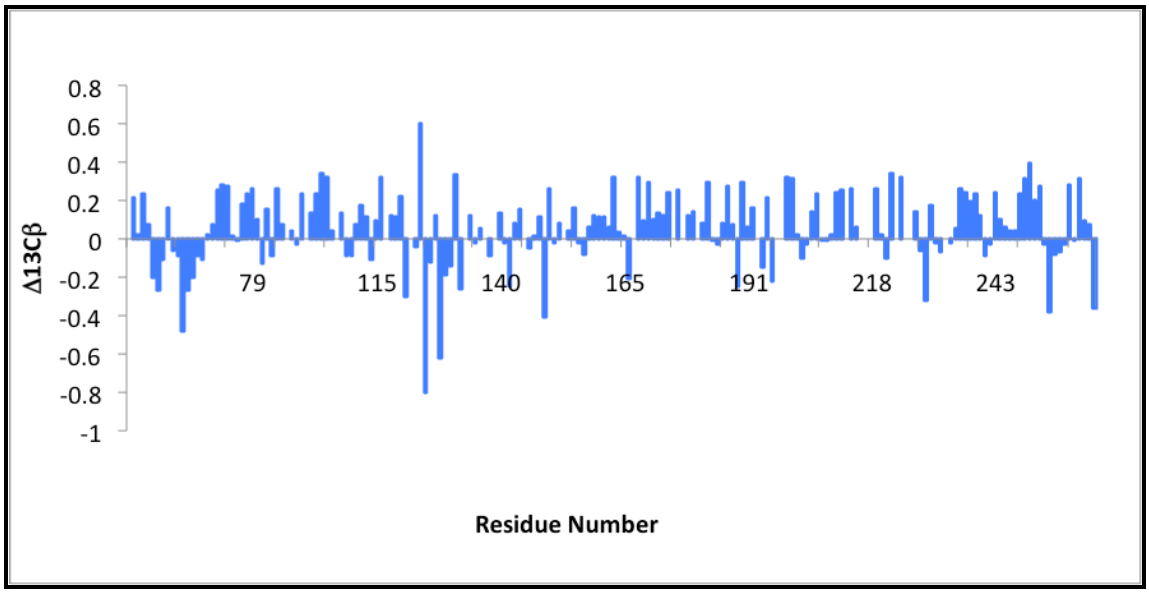


Figure 11: OPN220 ^{13}Cβ chemical shifts

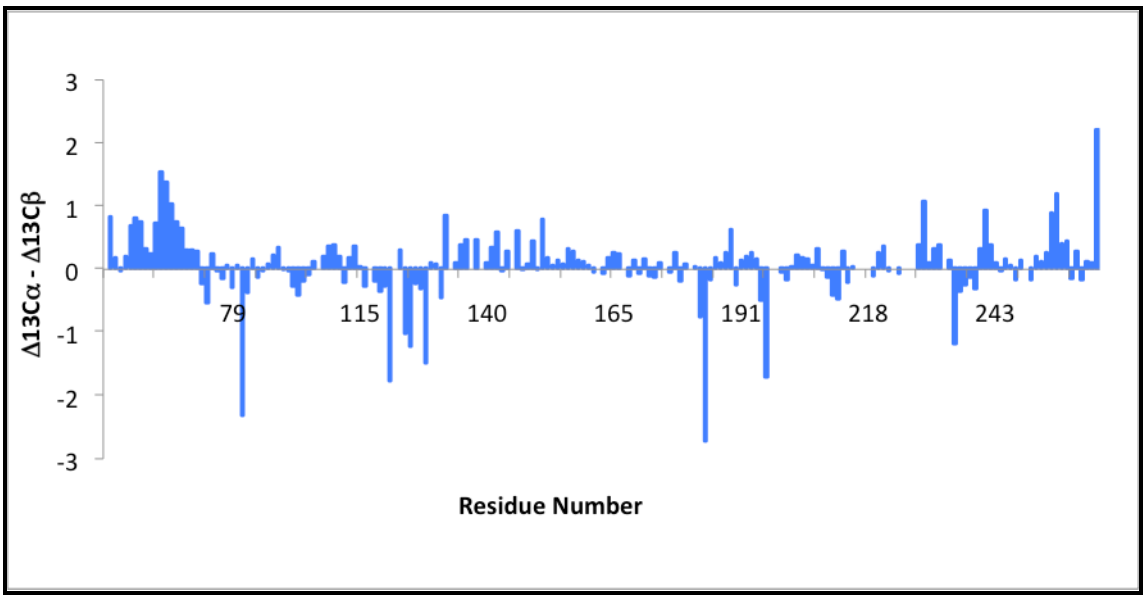


Figure 10 and Figure 11 represent the sequence corrected secondary ^{13}Cα, and ^{13}Cβ chemical shifts for OPN. In α-helices, ^{13}Cα appear to be shifted downfield relative to their random coil values, whereas ^{13}Cβ are shifted upfield, so the effect of both nuclei spins are opposite in sign. In β-strands the converse applies. Overall, the main parts of the residues display chemical shifts values typically for random coils. However, for some residues small secondary shifts indicating a slight weighting towards α-helical conformation are apparent (for example: Leu58 to Leu68 and the core region from 130-150).

The region 116-131 displays harbours a large number of proline residues. This feature combined with the presence of many residues with comparably large negative secondary chemical shifts hints to a tendency for formation of a polyproline helix in this region.

Residues 81, 184, 196 also display negative secondary chemical shift values, most probably due to the bordering prolines at these locations.

5.2 Relaxation Parameters of OPN

Measurement of NMR relaxation rates provides a window on protein dynamics over a broad range of time scales. The longitudinal ^{15}N-T_1 values, transverse ^{15}N-T_2 values, and heteronuclear steady state NOE $^{15}N\{^1H^N\}$ values are sensitive to dynamics on the pico-second to nanosecond time scale, whereas ^{15}N-T_2 can also be effected by conformational or chemical exchange processes on the millisecond to microsecond time scale. Despite the rather small chemical shift dispersion of osteopontin in the HQSC spectra, in total for 122 out of 183 assigned non-proline residues ^{15}N relaxation data could be obtained at 298 K and 600 MHz.

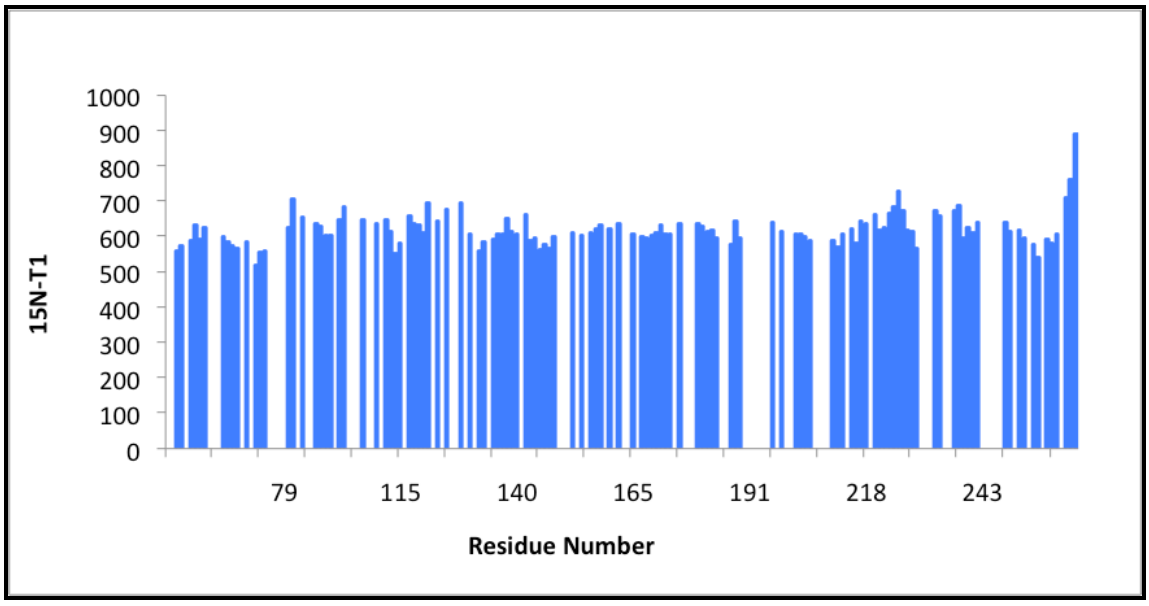


Figure 12: OPN220 ^{15}N-T_1 longitudinal relaxation times

As indicated in Figure 12 the ^{15}N-T_1 measurements exhibit a largely monotonous and featureless distribution along the amino acid sequence, except for the C-terminal end, with an average value of around 600 ms.

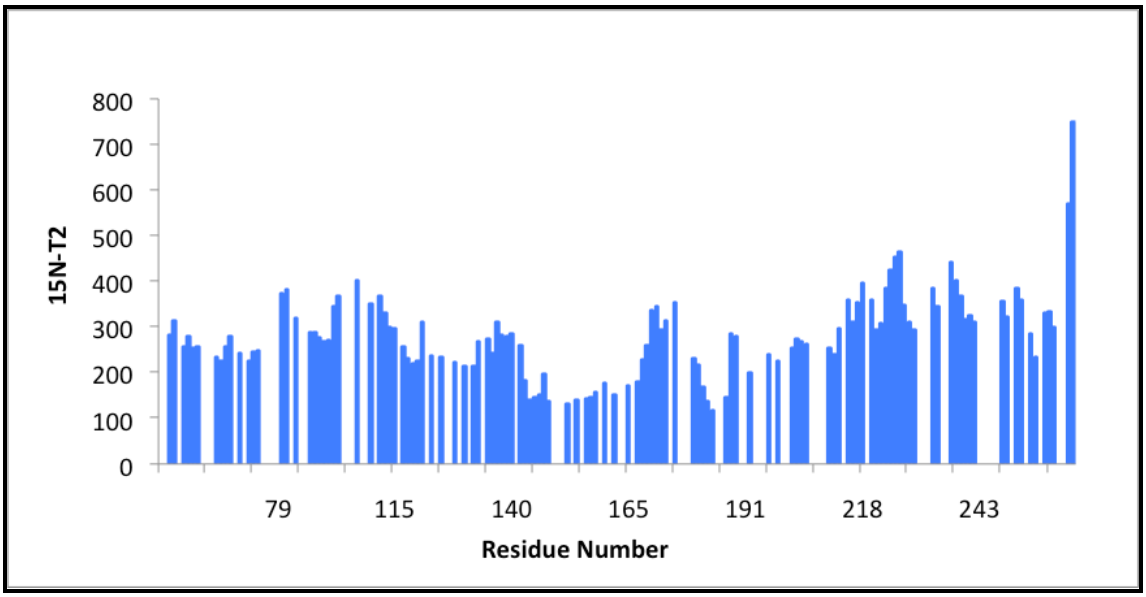


Figure 13: OPN220 ^{15}N-T_2 transversal relaxation times

The residue plot of the experimental ^{15}N-T_2 values (Figure 13) shows several sequential variations ranging from 114 ms to 747 ms, which is in stark contrast to an expected parabolic form (with ^{15}N-T_2 minimum in the centre of the sequence) arising from a complete random structure without any structural preferences. Overall there is a good agreement between secondary chemical shifts (2nd structure formation) and motional restrictions probed by ^{15}N T_2 relaxation. Interestingly, the stretch Leu182-Trp183-Trp184 also appears to be constrained presumably because of the spacious indol side chains.

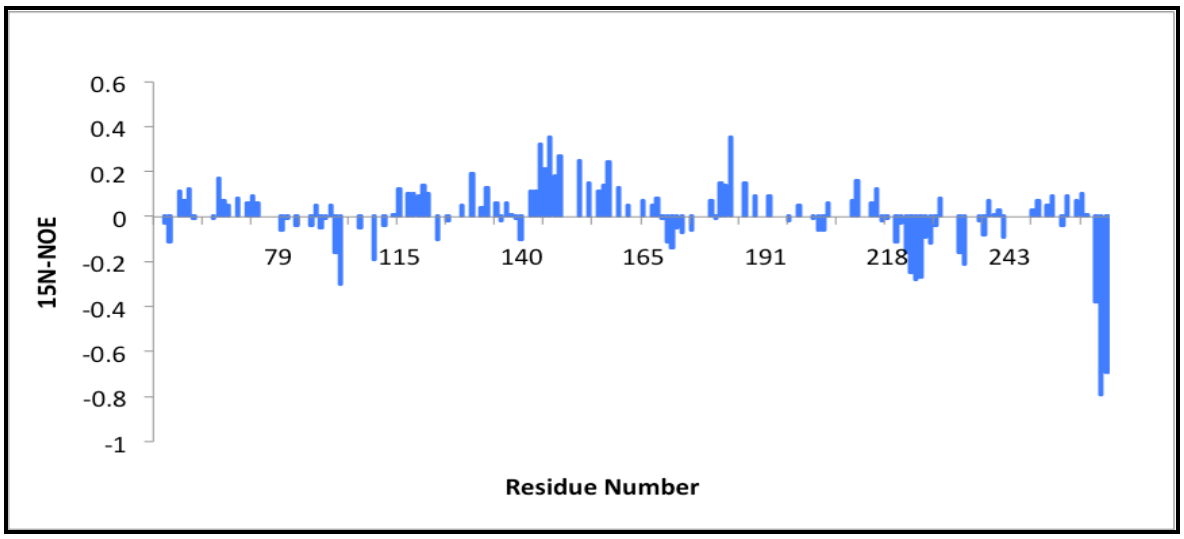


Figure 14: OPN220 heteronuclear ^{15}N - NOE

The obtained steady state NOE $^{15}N\{^{1}H^{N}\}$ (Figure 14) values in the center of the sequence are also found to be generally higher and strict positive in this area. Highest values (0.2 to 0.37) are observable in the middle of the sequence and for Trp184 (0.35). Extensive patches of negative NOE values are apparent from residues Ala170 to Gly175, and more pronounced along a 9-residue stretch comprising residues Gly217 to Gly225, a motif build up by residues of small side chains (3 glycines, 2 alanines, 2 serines). In addition, these parts are displaying higher $^{15}N\text{-}T_2$ values about 350 to 450 ms. The total average of $^{15}N\text{-}T_2$ is 280 ms. The intramolecular dynamics of osteopontin on a μs-to-ms time scale was investigated by state-of-the-art ^{15}N-CPMG dispersion experiments (see Material and Methods). Numerical analysis of the relaxation rate $R_{1\rho}$ *vs* CPMG-frequency (ν_{CPMG}) profile, however, revealed for all the residues (resolved in the 2D $^{15}N\text{-}^{1}H^{N}$ HSQC) flat dispersion profiles and thus indicating the lack of μs-to-ms time scale motions in OPN.

5.3 Residual Dipolar Couplings for OPN unfolded states

Another very useful tool for the characterization of IUPs involves the measurement of Residual Dipolar Couplings (RDCs). RDCs report on time-averaged conformations and can be used to understand both structure and dynamics of disordered proteins. RDCs are measured by weakly aligning a macromolecule in slightly anisotropic media, such as filamentous phages like PF1. The small degree of alignment resulting from the anisotropic environment leads to incomplete averaging of the dipolar couplings between magnetic nuclei close in space (typically directly attached). The magnitude of the RDC is dependent on the orientation of an internuclear vector such as $^{15}N\text{-}^{1}H$ relative to the alignment tensor of the protein as a whole. Therefore, valuable information about the relative orientation of individual bond vectors can be obtained.

The expected distribution of RDCs in unfolded proteins is globally followed in OPN (Figure 15), with negative couplings throughout the chain.

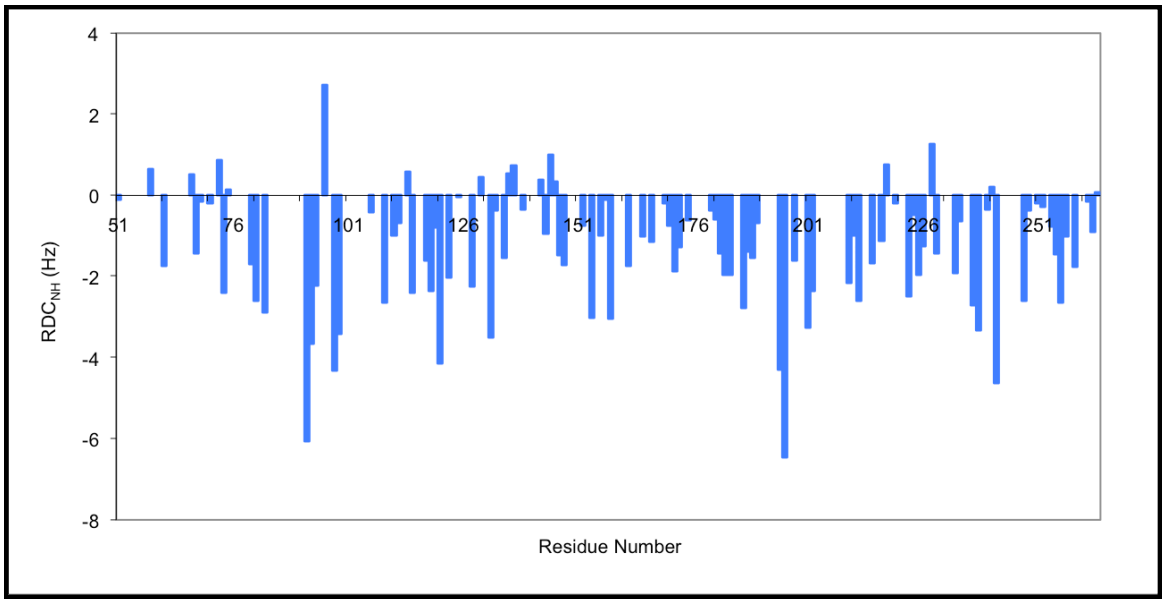


Figure 15: RDC-Profile of OPN220.

Large negative RDCs are associated with locally more extended conformations. In more extended parts of the chain the NH interaction direction will tend to be perpendicular to the external field giving rise to negative RDC. Interestingly, some stretches within the polypeptide chain show positive RDC values. The occurrence of such sign inversion in unfolded chains has been interpreted as an increase in local propensity to form turns or α-helical conformations.

Positive RDC values in the regions 58-73 and 130-146 nicely correspond to data from secondary chemical shifts and meta-structure prediction, which also show α-helical propensities for these areas. Areas with negative RDC's also compare with data from CD and secondary chemical shift, suggesting a large portion of the protein to occupy extended β-strand like conformation.

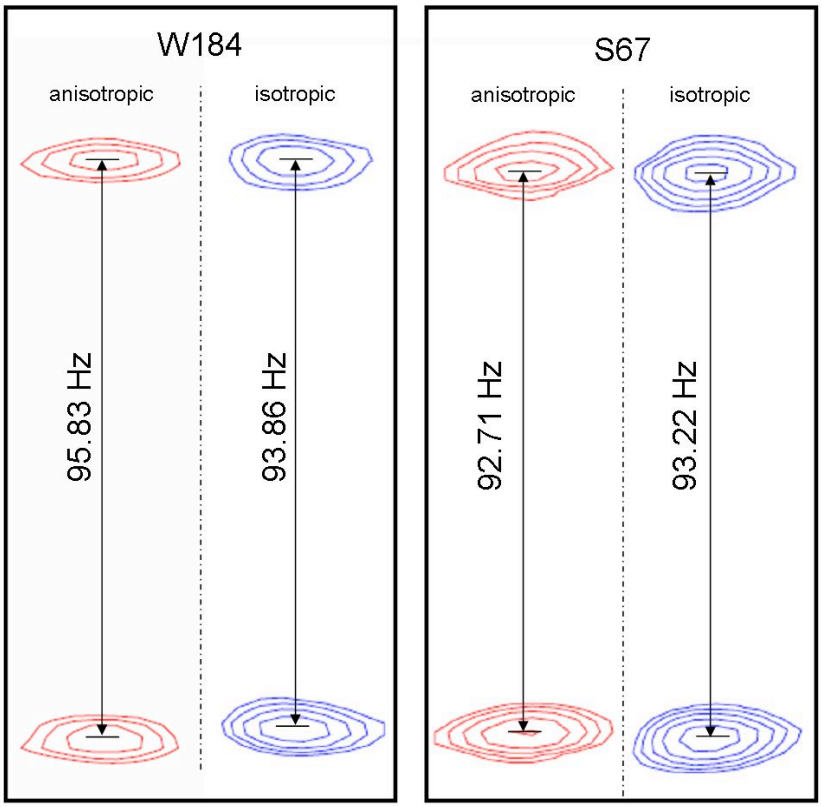


Figure 16: Isotropic and anisotropic doublets (aligned in PF1) of OPN220

5.4 Probing of Long-Range Interactions in the Native State of OPN with PRE.

While chemical shifts, and short and medium-range NOEs provide valuable insights into the secondary structural propensities of the polypeptide backbone in unfolded and partly folded proteins, it has generally proved difficult to observe long-range NOEs for unfolded or partly folded states. To circumvent this problem, site-specific nitroxide spin labeling can be utilized to probe transient long-range interactions in disordered proteins.

In an effort to evaluate the contribution of local and non-local interactions on the residual structure of OPN, site-directed spin labeling (SDSL) of three different residues of OPN was performed. Because the primary sequence of OPN lacks cysteine, three different cysteine-containing mutants (S108C, S193C and S247C) were constructed to provide attachment points for the nitroxide radical 4-(2-Iodoacetamido)-TEMPO (IAAT) (Figure 17).

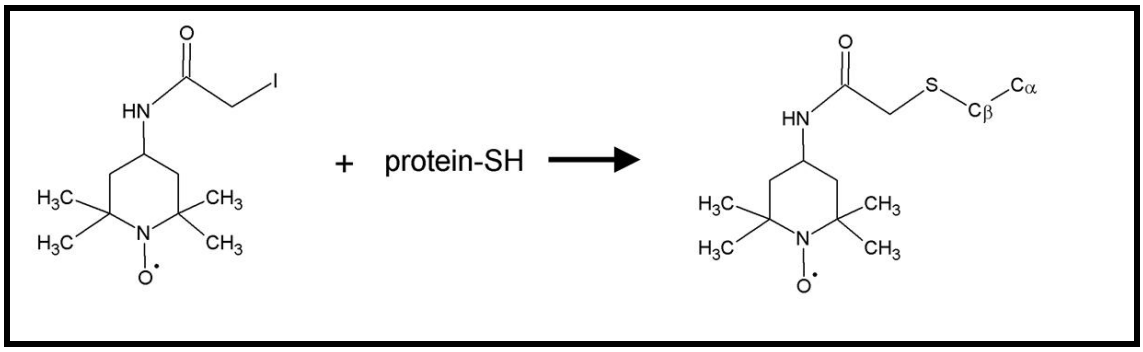


Figure 17: TEMPO attachment to cystein sidechains.

The target residues for the cysteine substitution where targeted to regions of low compactness (Figure 18), predicted with Robert Konrats meta-structure algorithm. Within these regions, basepair triplets translating to a cysteine by introducing a single basepair substitution were chosen:

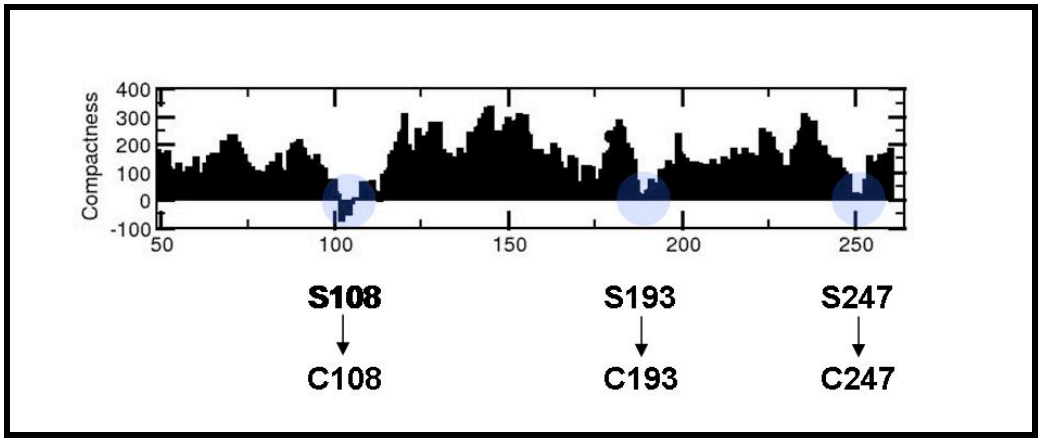


Figure 18: Compactness of OPN220

The interaction between a specifically attached paramagnetic nitroxide radical and nearby (< 25 Å) protons causes a broadening of their NMR signals because of an increase in transverse relaxation rate. This effect has an r^{-6} dependence on the electron–proton distance and thus allows the detection of long-range interactions in proteins. The peak intensity ratios between the two ^{15}N-^{1}H HSQC NMR spectra, i.e., in the presence and absence of the nitroxide radical (I_{param}/I_{diam}), permit the estimation of distances between the spin label and the affected amide protons in the protein.

The effects of the nitroxide radical on the NMR spectra were very different for the two analyzed cysteine positions. Intensity ratios for the S108C (Figure 19) mutant showed a significant paramagnetic effect extending to residue 167. The same effect is present in the S193C mutant (Figure 20) suggesting a rather compact structure in the core region (126-162) of OPN220. This region also comprises the helical patch from RDC data spanning the region from AS 130-146. This pattern is globally followed for S249C (Figure 21) although a larger scattering of intensity ratios was observed compared to the other two cystein mutants.

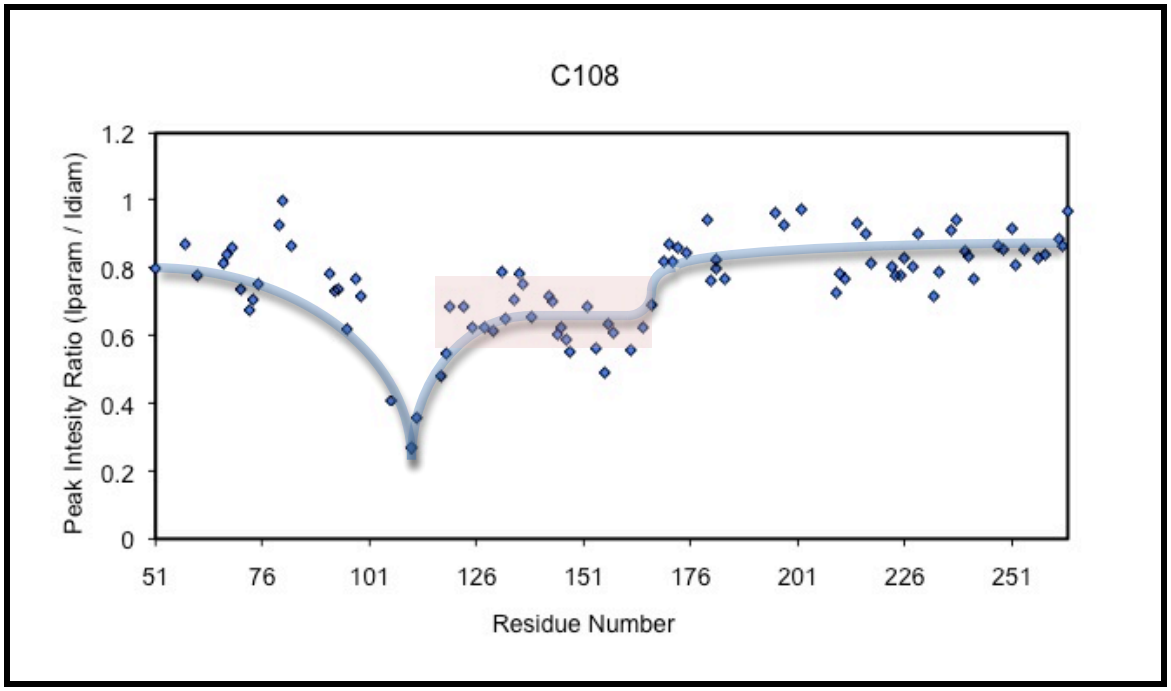


Figure 19: PRE –Profile of TEMPO attached to C108-OPN220 mutant.

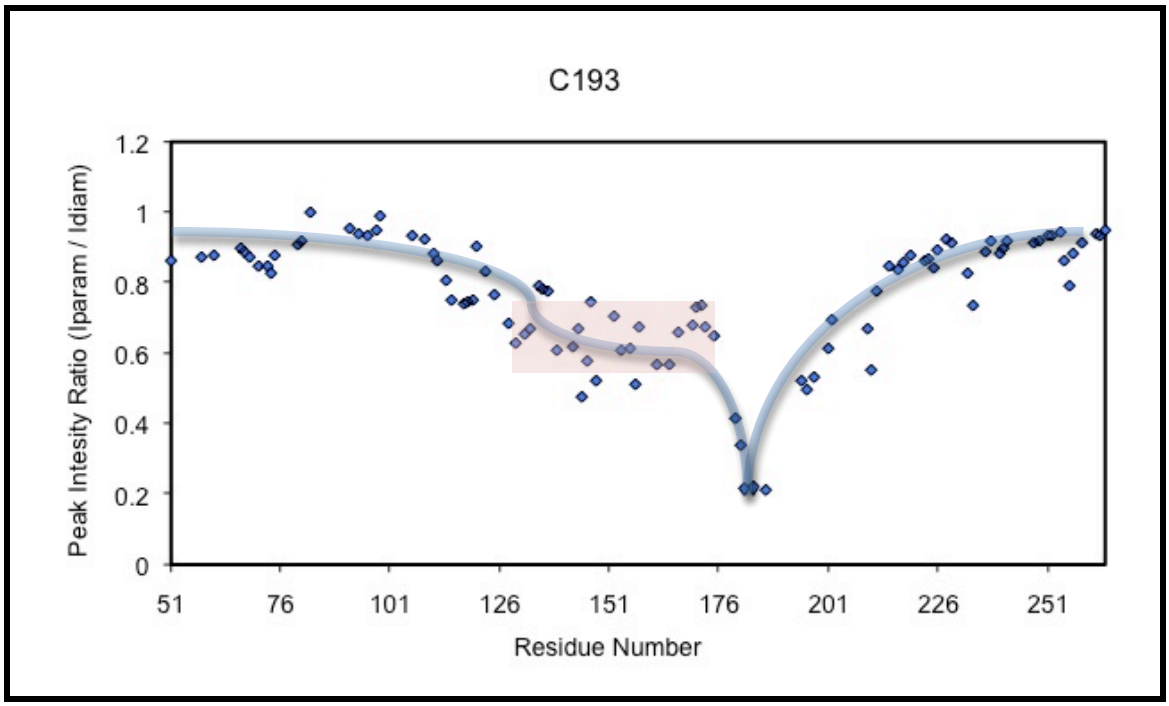


Figure 20: PRE –Profile of TEMPO attached to C193-OPN220 mutant.

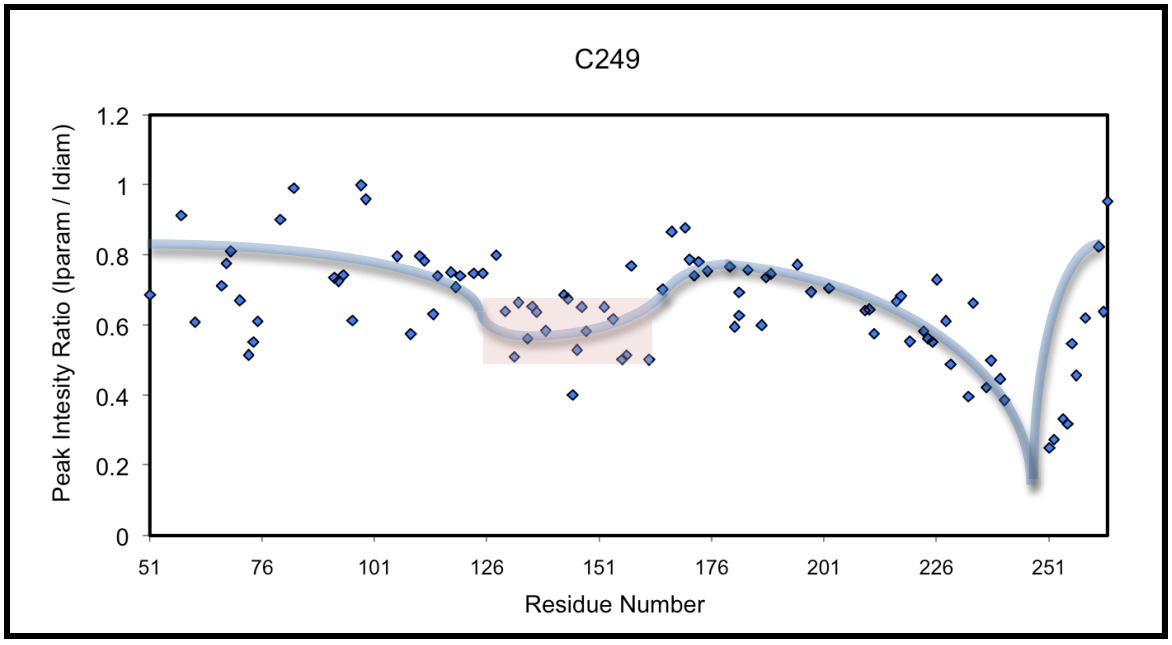


Figure 21: PRE –Profile of TEMPO attached to C249-OPN220 mutant.

5.5 Quantitative determination of secondary structure elements by Circular Dichroism (CD)

CD is a convenient and widely used method for studying the conformations and conformational changes of proteins in solution. Mean ellipticity θ for OPN220 (Figure 22) was measured at 20°C at 0.5 nm intervals from 180 – 260nm:

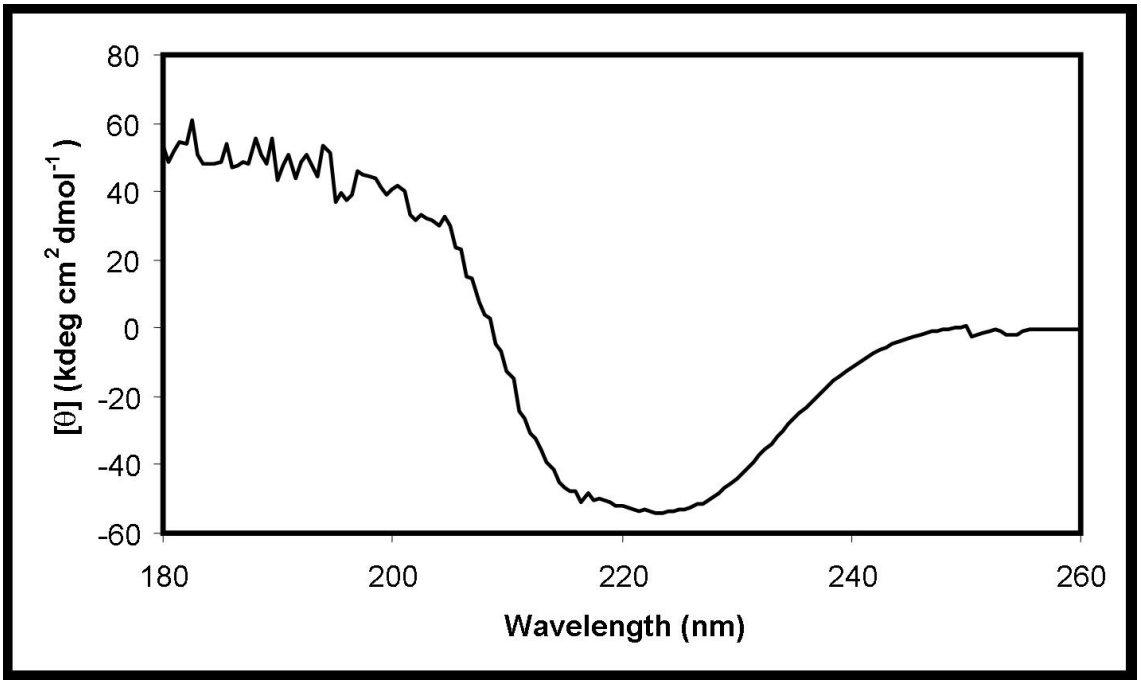


Figure 22: Mean ellipticity θ for OPN220

Deconvolution of the OPN spectra was achieved by comparing it with 33 reference base spectra thereby calculating the contributions of the various components to the secondary structure:

	180-260 nm	185-260 nm	190-260 nm	195-260 nm	200-260 nm	205-260 nm	210-260 nm
Helix	15.00%	9.50%	9.70%	10.60%	11.00%	7.70%	7.90%
Antiparallel	52.90%	55.40%	46.40%	35.00%	33.20%	33.20%	33.80%
Parallel	2.80%	4.20%	4.60%	5.80%	5.90%	5.50%	5.50%
Beta-Turn	16.70%	16.90%	16.50%	17.60%	17.50%	18.80%	19.30%
Rndm. Coil	24.60%	25.40%	28.40%	33.50%	34.70%	34.90%	35.10%
Total Sum	112.00%	111.40%	105.70%	102.60%	102.30%	100.10%	101.50%

The data obtained by CD strongly support data from RDC and secondary chemical shifts on OPN. Extended beta-strand like structures, mainly of antiparallel nature, show the highest contribtion to the secondary structure ensemble (30-50%). In addition, random coil contribution varies between 25-35%, supporting the idea of a largely unstructured conformation of Osteopontin.

The total helical amount varies between 8-15% depending on the wavelength range. This is in very good agreement with RDC and secondary chemical shift data suggesting α-helical propensities in the regions 58-73 and 130-146.

CD thermal denaturation studies for Osteopontin were conducted at 218 nm (Figure 23) with no obvious melting behaviour detectable in the range from 20-60°C. This corroborates the lack of a stable tertiary structure of OPN.

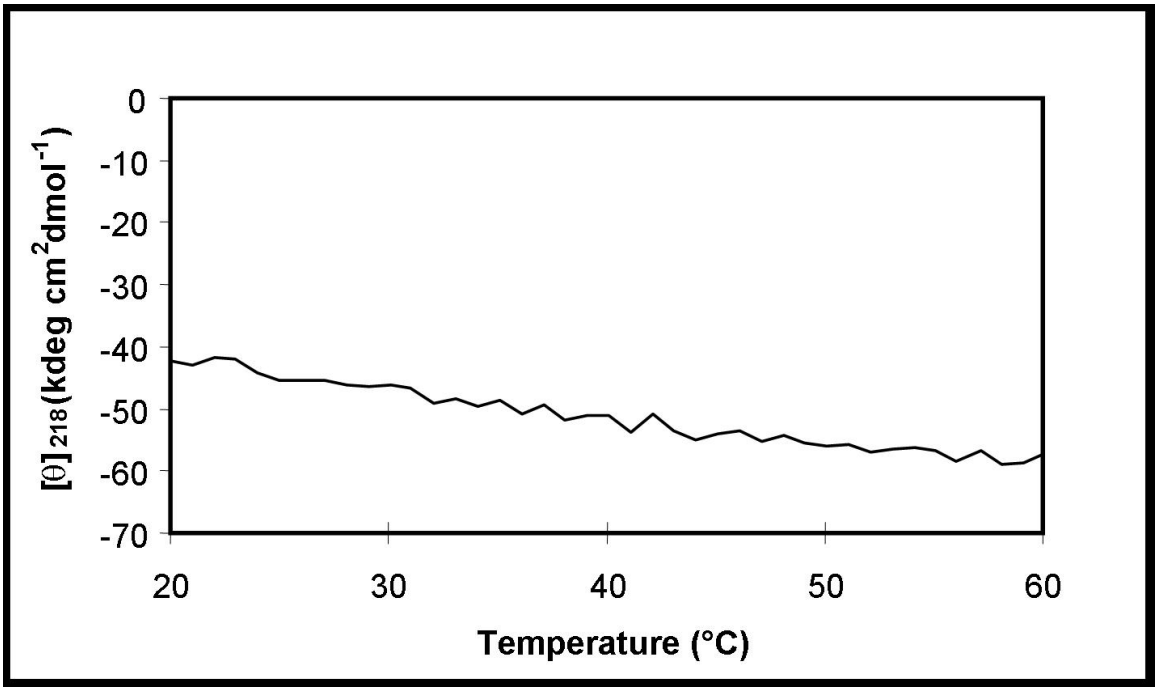


Figure 23: Thermal denaturation curve for OPN220

5.6 Hydrodynamic radius measurement

A pulsed field gradient (PFG) NMR diffusion measurement was conducted to get a rough idea on the diameter of the molecule in solution. The hydrodynamic radius (R_h) of OPN

was measured to 41.07 Å. According to the empirical relationship between the length of a polypeptide chain and its hydrodynamic radius [57], a globular 220 residues 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 relatively large R_h of OPN (41.07 Å) in comparison to expected value for a globular form (22,7 Å) indicates that under these conditions OPN is mostly unfolded. Nevertheless, the lower R_h of OPN compared to the expected value for a completely unfolded chain (47.8 Å) suggests that OPN retains some local compactness in solution or samples conformations in which the polypeptide is more compact (presumably due to transient secondary elements as well as long-range residue contacts being formed).

Independent experimental data for the hydrodynamic behaviour of OPN were provided by a small-angle X-Ray scattering analysis. The one-dimensional scattering curves were further treated using the program GIFT. The interaction of the proteins in solution was negligible since the concentration was low and any charge-contribution from ions were excluded by subtracting a buffer-only spectra from the protein-buffer spectra. The scattering was thus only related to the overall shape, the electron density and its distribution within the protein. The pair distance distribution function (PDDF) was obtained by Indirect Fourier Transformation (IFT). This function can be seen as a histogram of distances occurring in the protein. The shape of the PDDF can be evaluated in order to gain information about the structure of the protein. A Gaussian coil of a polymer has a linear increase of the PDDF at low distances. At higher distances the function shows a slow abatement. Figure 24 shows the PDDF extracted from the experimental scattering curve of OPN.

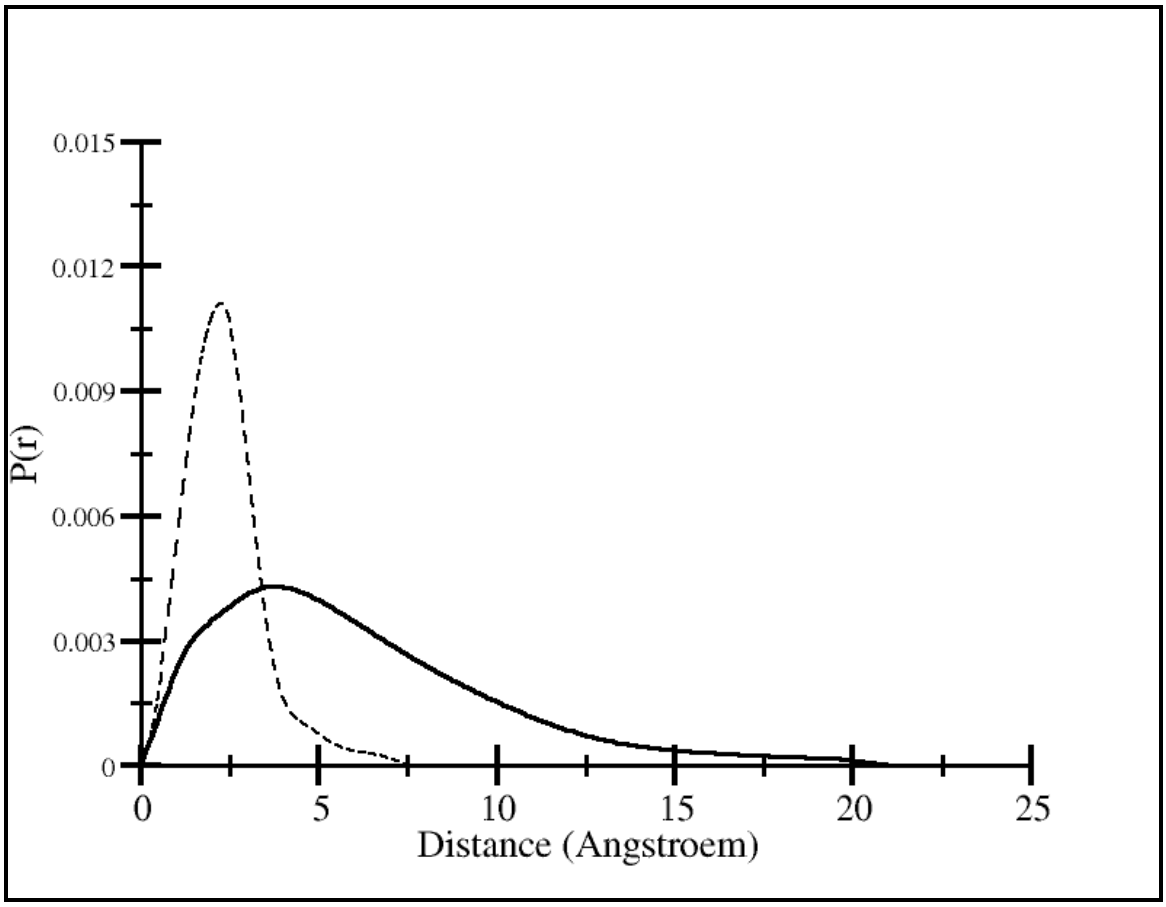


Figure 24: Pair Distance Distribution Function (PDDF) for Osteopontin (solid line) and Q83 (dashed line)

The linear increase can clearly be observed in the PDDF of the protein. The decrease of the function is flatter than it would be expected for a Gaussian coil. This indicates that the structure of the protein is more elongated than a conventional Gaussian-coil. The maximum dimension of the protein can be read off of the PDDF and it is about 210 Å. Furthermore a shoulder in the function at about 15 Å can be observed. It is important to mention at this point that the scattering curve is a sum of the scattering, which is caused by all structures appearing in the measured volume during the measurement time. So the extracted PDDF contains contributions of all structures occurring in the sample. An explanation for the shoulder in the PDDF provides evidence for the fact that the protein exists in more than one structure. For comparison, **Error! Reference source not found.** also shows the PDDF curve of the stably folded lipocalin protein Q83. The strikingly different hydrodynamic behavior is clearly identified by the SAXS data. The average

radius of gyration of Q83 was determined to 21.3 Å. That is in very good agreement with the 3D structural data obtained by NMR spectroscopy.

5.7 Ligand Interaction

OPN was reported to bind Heparin by several articles although no experimental evidence was given. [58][59]. In order to provide more specific information about the binding site and the mode of binding we performed ^{15}N-^{1}H HSQC spectroscopy.

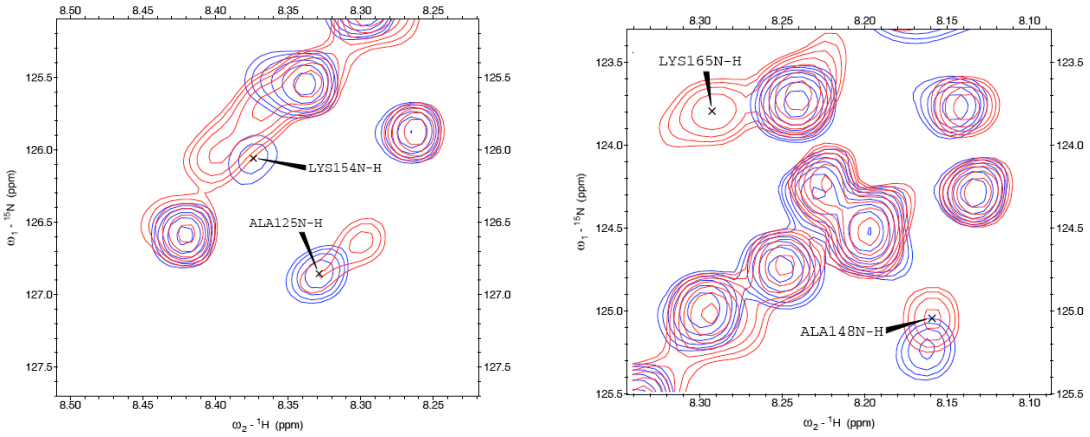


Figure 25: Overlay of HSQC spectra of free OPN220 and when bound to Heparin

Figure 25 shows an overlay of HSQC spectra of unligated OPN and when bound to Heparin. As can be seen from the figure there is evidence for 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 magnitude of chemical shift changes were small, presumably because Heparin lacks aromatic ring systems and the fact that ^{15}N-^{1}H HSQC spectra report on the protein backbone and not on side-chains which are in direct contact with the ligand. Mapping the residues, which display frequency changes upon binding of Heparin on the protein backbone, clearly reveals a distinct binding side near the central Integrin-binding domain (residues 117 to 130 and 142 to 167). This is in striking contrast to existing domain analysis of OPN where the Heparin binding domain was located to the C-terminal domain (between Asp298-Ile305). Interestingly, Heparin-binding inhibited Thrombin-

induced cleavage of OPN (data not shown) as the Heparin interaction site is overlapping with the RSK recognition sequence for thrombin cleavage.

Additionally, calcium was also reported to bind to OPN, and the putative Calcium binding site was assigned to the residue stretch Asp216-Ser228. Again, ^{15}N-^{1}H HSQC spectroscopy was performed. However, no evidence of calcium binding could be found. The ^{15}N-^{1}H HSQC spectra were nearly unchanged even at a 1:30 molar access of calcium over OPN. This is in agreement with recently published data stressing the importance of phosphorylation for OPNs role in biomineralization processes [38]. Recombinant, and therefore nonphosphorylated OPN is not capable of binding calcium.

6 Discussion

Structural and dynamic properties of Osteopontin (OPN) were investigated by means of multi-dimensional NMR spectroscopy and ^{15}N relaxation parameters with supporting data derived from SAXS and Circular Dichroism (CD) measurements. Global parameters of the protein including compactness and composition of secondary structure elements are a good basis for further and more detailed investigations of a protein. Data from CD measurements already indicated a certain degree of secondary structure elements present in Osteopontin although the CD-data suggests that most of the protein seems to be in an extended β-strand like and/or random coil conformation. These findings are further supported by significant ^{13}C secondary shifts observed for various segments along the polypeptide chain also hinting to the existence of local 2nd structure elements.

The correlation between deviations from random coil chemical shift values and backbone dynamics provides convincing evidence for 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 (random coil like) and rapidly interconverting α-helical conformations but rather exists in cooperatively folded local α-helices. Interestingly, these preformed and motionally restricted structural motifs comprise the biochemically established protein interaction motifs of OPN. No evidence for slower (ms-to-ms) time scale motions was found (based on ^{15}N-CMPG dispersion measurements).

The compactness of the protein is another interesting global parameter.

A pulsed field gradient (PFG) NMR diffusion measurement yielded a hydrodynamic radius (R_h) for OPN of 41.07 Å. This result clearly differs from R_{h220} expected for a globular (R_{h220} ~23 Å) and for a completely unfolded protein (R_{h220} ~48 Å). OPN is therefore neither a stably folded protein, nor completely unfolded with no apparent structure at all. Obviously OPN retains some local compactness in solution due to secondary elements present in the proteins structural ensemble. These findings were also supported by SAXS data recorded in Graz showing a maximal internal protein distance of

210 Å, although the majority of the protein ensemble is present in a more compact conformation.

Information about compactness and secondary structure composition can be extracted out of the primary protein sequence using the meta-structure approach introduced by Robert Konrat. The meta-tructure approach allows sequence-based calculation of residue-specific compactness and local 2nd structural features in proteins. [56] 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 (e.g. inversely related to surface exposure). The combination of residue compactness and local 2nd structure as a function of residue position can be regarded as the protein's meta-structure. Large values of C_i are found for residues located in stable parts of the protein and on average deeply buried in the interior of the structure. In contrast, small values are found for flexible loop regions and/or intrinsically unfolded segments of the polypeptide chain.

A survey of PDB structures revealed an average compactness value of about 300 for a typically folded protein. The (protein specific) average compactness value can thus be used as a (coarse) benchmark value for screening for protein foldedness. The predicted average compactness values of OPN amounts to about 182. This is clearly below the average compactness value of 300 for all PDB entries. OPNs compactness is clearly below values observed for stably folded proteins, and thus supporting the notion that OPN is an intrinsically unstructured protein like the other members of the SIBLING family.

Overall the NMR derived date is in good agreement with predicted meta-structure data. Increased compactness values were found for the region spanning residues 110 to 190 aswell as 200 to 250. The region ranging from 110 to 190 shelters several binding sites for various integrins (residues 130-150). ^{15}N-T_2 data shows a motional restriction in exactly this region. ^{15}N-NOE data further supports these findings showing generally higher and strictly positive ^{15}N-NOE values in this region. The motional restriction hints to a preformed structure, although not highly populated, being formed in this region. This could be a necessary perquisite for OPN in order to interact with the various integrin isoforms. Interestingly, RDC values in this region are found to be positive, indicative of a

α-helical conformations, fully in agreement with secondary chemical shifts values in this region of the protein. It can be hypothezised that the interaction mode of integrins with osteopontin very likely requires OPN to adapt a defined secondary structure allowing it to interact with an integrin receptor.

Another site of interest within osteopontin's primary sequence is its putative CD44 interaction site located n the C-terminal part of the protein (in proximity of residue 220). NOE values in this region are strictly negative and no motional restriction is apparent from ^{15}N-T_2 data in this region. The interaction mode for OPN-CD44 is very likely different from that of OPN-integrin. Since the structure of the OPN in this region doesn't seem to be motionally restricted, an interaction mode similar to p27Kip1 bound to the cyclin A–cyclin dependent kinase 2 (Cdk2) complex (Figure 9) could be occurring. Since CD44 is a very large protein, OPN might retain it's open conformation interacting with the CD44 via a large binding surface, thereby retaining most of its open conformation. This is a common mode of interaction for IUP's, completely different from the second mode, requiring a preformed secondary structure to engage it's binding partner as is the case for integrin binding.

Taking the results of this diploma thesis together, OPN can be regarded as a flexible polypeptide chain housing local structural and functional interaction motifs separated by flexible linker regions. This modular architecture and intrinsic flexibility may be a necessary prerequisite for osteopontin's biological functionality as it allows for substantial structural plasticity necessary to accommodate various protein binding partners recruited to the cellular surface.

7 Materials and Methods

7.1 Microorganisms

Organism	Strain
Escherichia (E.) coli	DH5 α
	BL21(DE3)

7.2 Media

LB

20 g LB-Broth

dissolve in 1l H_2O.
autoclave at 121°C
store at 4°C

LB selective media

add 1 ml of ampicillin (100 mg/ml) and/or chloramphenicol (25 mg/ml) and/or kanamycin (25 mg/ml) and/or tetracycline (5 mg/ml) to 1l LB
prepare fresh

LB selective agar

500 ml LB
7.5 g agar agar
autoclave at 121°C for 15 min, cool to 55°C,
add 500 µl of appropriate antibiotics, pour into plates
store at 4°C

M9 media

6 g $Na_2HPO_4x2H_2O$
3 g KH_2PO_4
0.5 g NaCl
1 g NH_4Cl

dissolve in 1 l H_2O, autoclave at 121°C for 15 min
add:
20 ml of 20% glucose
2 ml of 1 M $MgSO_4$
10 ml trace elements
0.2 ml 1 M $CaCl_2$

M9 media

for ^{15}N-labelling of proteins

see M9 media
use $^{15}NH_4Cl$ instead of NH_4Cl

M9 media

for ^{15}N, ^{13}C-labelling of proteins

see M9 media
use $^{15}NH_4Cl$ instead of NH_4Cl
use 2 to 4 g ^{13}C-Glucose instead of Glucose (add after autoclaving)

NZCYM medium

21.98 g NZCYM
add H_2O to a final volume of 1 l

autoclave at 121°C for 15 min.
store at 4°C

TY 2x

16 g tryptone
10 g yeast extract
5 g NaCl
add H_2O to a final volume of 1 l

autoclave at 121°C for 15 min.
store at 4°C

7.2 Buffers & Solutions

Seperation Gel

33% (w/v) acrylamide
0.9% (w/v) bisacrylamide

33g acrylamide
0.9 g bisacrylamide (N,N'-methylene bisacrylamide)
add H_2O to a final volume of 100 ml
sterilize by filtration (0.2 µm)
store light-protected at 4°C

Stacking Gel

20% (w/v) acrylamide
1% (w/v) bisacrylamide

20 g acrylamide
1 g bisacrylamide (N,N'-methylene bisacrylamide)
add H_2O to a final volume of 100 ml
sterilize by filtration (0.2 µm)
store light-protected at 4°C

Ammoniumperoxodisulfate (APS)

10% (w/v)

1 g ammoniumperoxodisulfate
add H_2O to a final volume of 10 ml
store at -20°C

Ampicillin

100 mg/ml (≅0.27 M)

1 g ampicillin
add H_2O to a final volume of 10 ml

sterilize by filtration (0.2 µm)
store at -20°C

$CaCl_2$

1M

14.7g $CaCl_2 \times 2H_2O$
add H_2O to a final volume of 100 ml
autoclave at 121°C for 15 min

Chloramphenicol

25 mg/ml

add EtOH to a final volume of 10 ml
store at -20°C

Coomassie staining solution

0.5% (w/v) Coomassie
50% (v/v) EtOH
10% (v/v) HOAc

2.5 g Coomassie Brilliant Blue R-250
250 ml EtOH (98%)
50 ml HOAc (99%)
add H_2O to a final volume of 500 ml
store light protected

Destaining solution

30% (v/v) EtOH
10% (v/v) HOAc

300 ml EtOH (98%)
100 ml HOAc (99%)
add H_2O to a final volume of 1 l

Dithiothreitol (DTT)

1 M

1.54 g dithiothreitol
add H_2O to a final volume of 10 ml
sterilize by filtration (0.2 µm)
store at -20°C

DNA sample buffer (5x)

20 mM EDTA
30 % (w/v) Glycerine
0.5 % (w/v) SDS
0.1 % (w/v) Bromphenol blue

100 µl EDTA 1 M (pH 8.0)
5 mg bromphenol blue
1.7 ml glycerine
125 µl SDS 20 %
add H_2O to a final volume of 5 ml
store at 4°C

dNTP Mix

2 mM dATP
2 mM dCTP
2 mM dGTP
2 mM dTTP

10 µl dATP 100 mM
10 µl dCTP 100 mM
10 µl dGTP 100 mM
10 µl dTTP 100 mM
add H_2O to a final volume of 500 µl
aliquote and store at -20°C

HCl

1 M

100 ml HCl_{conc} (32 %, ρ=1.16)
add H_2O to a final volume of 1 l

IPTG

1 M

2.4 g IPTG (238.25 g/mol)
add H_2O to a final volume of 10 ml
sterilize by filtration (0.2 µm)
store at -20°C

Kanamycin

25 mg/ml

250 mg Kanamycin monosulfate
add H_2O to a final volume of 10 ml
sterilize by filtration (0.2 µm)
store at -20°C

Laemmli buffer

25 mM Tris
190 mM Glycine
0.1 % SDS

6 g tris base
28.75 g glycine
2 g SDS
add H_2O to a final volume of 2 l

$MgCl_2$

1 M

20.33 g $MgCl_2 \times 6H_2O$
add H_2O to a final volume of 100 ml
autoclave at 121°C for 15 min

NaCl

1 M

58.44 g NaCl
add H_2O to a final volume of 1 l
autoclave at 121°C for 15 min

NaOH

1 M

40 g NaOH
add H_2O to a final volume of 500 ml

NaP buffer

100 mM sodium phosphate pH 6.5

adjust pH by mixing
100 mM $NaH_2PO_4 \cdot H_2O$ 150 mM NaCl
100 mM $Na_2HPO_4 \cdot 2H_2O$ 150 mM NaCl
to a final volume of 1 l

PBS buffer

140 mM NaCl
2.7 mM KCl
10 mM Na_2HPO_4
10 mM

8.18 g NaCl
0.2 g KCl
1.78 g Na_2HPO_4
0.24 g KH_2PO_4
add H_2O to a final volume of 1 l
autoclave at 121°C for 15 min

PBS buffer (high salt)

1 M NaCl
2.7 mM KCl
10 mM Na_2HPO_4
10 mM

58.44 g NaCl
0.2 g KCl
1.78 g Na_2HPO_4
0.24 g KH_2PO_4
add H_2O to a final volume of 1 l
autoclave at 121°C for 15 min

Protease inhibitor solution

2x Protein sample buffer

120 mM Tris×HCl, pH 6.8
6 % (w/v) SDS
20 % (v/v) Glycerine
0.01 % (w/v) Bromphenol blue
10 % (v/v) β-Mercaptoethanol

1.2 ml Tris×HCl 1 M (pH 6.8)
3 ml SDS 20 %
2 ml glycerine
1 mg Bromphenol blue
add H_2O to a final volume of 9 ml
store in 900µl aliquots at -20°C
thaw only once
add 100 µl of β-mercaptoethanol to 900 µl of buffer
store at 4°C

SDS 20%

20% SDS (w/v)

20 g SDS
add H_2O to a final volume of 100 ml

Stop Mix (5x)

20 mM EDTA
30 % Glycerine
0.5 % SDS
0.1 % bromphenol blue

0.1 ml 1 M EDTA (pH 8.0)
1.7 ml glycerine
0.125 ml 20 % SDS
2 ml 0.5% bromphenol blue
add H_2O to a final volume of 5 ml

TBE buffer (10×)

pH 8.3
0.89 M Tris
0.89 M Boric acid
0.02 M EDTA

108 g tris base
55 g H_3BO_3
7.44 g EDTA $Na_2 \cdot 2H_2O$
add H_2O to a final volume of 1 l
autoclave at 121°C for 15 min

Tetracycline

5 mg/ml

50 mg tetracycline
add H_2O to a final volume of 10 ml
sterilize by filtration (0.2 µm)
store at -20°C

7.3 DNA Work

7.3.1 Starter cultures for plasmid preparations

Inoculate 20 ml of LB medium supplemented with the appropriate antibiotics with a single colony from a transformation plate using a sterile pipette tip. Alternatively, use 1 µl of a glycerol stock for inoculation. Incubate at 37°C and 225 rpm to an OD_{600} of approximately 1.

7.3.2 Preparation of plasmid DNA from bacteria

Plasmid DNA from bacteria was obtained using Plasmid Mini or Midi Kits (Qiagen) according to the manufacturer's instructions.

7.3.3 Preparation of competent bacteria

- Prepare an overnight culture of the respective E. coli strain; inoculate NZCYM medium supplemented with the appropriate antibiotics with cells from the bacterial stock solution.
- Inoculate 100 ml of 2x TY medium with 1 ml of the overnight culture. Grow the cells at 37 °C and 225 rpm until the optical density at 600 nm (OD_{600}) reaches 0.5 to 0.6.
- Pour the culture into ice-cold 50 ml conical tubes and chill on ice for 10 min.
- Centrifuge at 4 °C for 10 min at 4000 x g. Discard the supernatant and resuspend the cells in 100 ml ice-cold TFB I solution. Incubate on ice for 10 min.
- Centrifuge as above and suspend the pellet in 4 ml of ice-cold TFB II solution.
- Immediately prepare 100 µl aliquots in pre-chilled 2 ml tubes and store them at –80 °C.
- Test the transformation efficiency by transforming 10 ng of pUC19 plasmid Dna according to the heat-shock protocol. The efficiency should be about $1x10^6$ to $1x10^7$ pfu per µg of DNA.

7.3.4 Transformation of chemically competent *E.coli*

- 100µl of chemically competent E.coli strains DH5 α or BL21(DE3) were thawn on ice and mixed with 1-2µl of plasmid (5-10 ng/µl).
- Following incubation on ice for 30 minutes, cells were kept at 42°C for 1.5 minutes and then put on ice for 2 minutes.
- After addition of 1ml of LB medium, cells were incubated at 37° under constant shaking for 1 hour.
- Cells were concentrated by short centrifugation at 13000 rpm and the supernatant was discarded, leaving approximately 100µl. The cell pellet was resuspended and plated on an agar plate containing the appropriate antibiotics
- Cells were incubated overnight at 37°C.

7.3.5 DNA Gel-Electrophoresis

Standard agarose gels for DNA analysis were prepared with 1% agarose in 1x TBE buffer. For fragments less than 300bp in length, the amount of

agarose was increased to 2%. The agarose gels were run at 40-60 V for 1 hr.

7.3.6 DNA Gel Extraction

The bands containing the DNA fragments of interest were cut out of the gel and DNA was eluted with the QIAEXII Gel Extraction Kit (Qiagen) according to the manufacturer's protocols.

7.3.7 Cloning of DNA fragments

Digestion of DNA with restriction endonucleases

- Digest DNA in a total volume of 10 µl per µg of plasmid.
- Add DNA template, about 0.1 µg/µl.
- Add 1/10 of 10x reaction buffer.
- Adjust the total reaction volume with H_2O, minus the volume of restriction enzyme.
- Add 1 to 2 units of restriction enzyme(s) per µg of DNA template; the volume of the enzyme added must not exceed more than 10% of the total volume, as the glycerin within the enzyme buffer may inhibit restriction.
- Incubate at the recommended temperature for 2 hrs up to o/n.
- If necessary terminate the reaction by incubation of the solution at the inactivation temperature for 10 to 20 min or by adding 1/4 of the total volume of 5x DNA sample buffer.
- Store at 4°C or directly load onto a gel.
- If the DNA has to be cleaved with two restriction enzymes or modified with other enzymes e.g. alkaline phosphatase or blunt ending enzymes, the reactions can be carried out simultaneously if both enzymes are active in the same buffer. Alternatively, the enzyme with the buffer of lower ionic strength should be used first. The appropriate amount of salt and the second enzyme can be added subsequently. Another way is to precipitate DNA after the first digestion and then perform the second one.

Ligation of DNA fragments

- In an Eppendorf tube mix:

Vector DNA	100 ng	1 µl
DNA insert	100 ng	3 µl

10x T4 DNA Ligase buffer		1 µl
T4 DNA Ligase		1 µl
H_2O	to 10 µl	4µl
		10 µl

- incubate at 22 °C for 1.5 hrs.
- inactivate DNA Ligase at 65 °C for 10 min.

7.3.8 Quantification of nucleic acids

- Prepare 100 µl dilutions of DNA in H_2O in concentrations of about 5 to 50 µg/ml.
- Measure the absorbance at 260 and 280 nm in a UV Spectrophotometer.
- The following assessments can be used for quantification:
  DNA: A_{260} = 1 at a concentration of 50 µg/ml
  RNA: A_{260} = 1 at a concentration of 40 µg/ml
- A ratio of A_{260}/A_{280} between 1.6 and 2.0 indicates pure DNA or RNA without significant protein contamination.

7.3.9 Polymerase Chain Reaction (PCR)

Site-directed mutagenesis using cycle PCR

This protocol is useful to introduce point mutations, replace amino acids, and to delete or insert single or multiple adjacent amino acids. It is based on the QuikChange II Site-Directed Mutagenesis Kit (Stratagene).

- Prepare two mutagenic primers that contain the desired mutation and anneal to the same sequence on opposite strands of the plasmids.
- Mutation should be located in the center of the primer sequence, flanked by 10 to 15 correct bases on each side.

- In a PCR tube mix:

DNA template	50 ng	2 µl
oligonucleotide primer #1	100 ng	5 µl
oligonucleotide primer #2	100 ng	5 µl
dNTPmix		1 µl
10x PCR buffer		5 µl
H_2O	to 50 µl	31µl
PfuUltra HF DNA polymerase	2.5 U/µl	1 µl
		50 µl

- PCR program:

denaturation		95 °C	00:00:30
15x	denaturation	95 °C	00:00:30
	annealing	55 °C	00:01:00
	elongation	68 °C	1 min
final elongation		68 °C	00:10:00
hold		4 °C	∞

- Add 1 µl of DpnI (10U/ µl) to the amplification reaction and incubate at 37 °C for 1 hour. DpnI digests the parental supercoiled dsDNA.
- Transform 1 µl of the digested PCR product into appropriate cells and plate on agar plates containing the appropriate antibiotics.

7.4 Protein Work

7.4.1 Starter cultures for recombinant protein expression

Inoculate 20 ml of LB media and the appropriate antibiotics with one colony from a transformation plate.
Incubate at 37 °C and 180 rpm until OD_{600} reaches ~ 1.0.

7.4.2 Expression of recombinant proteins

- Add 10ml of the starter culture to the expression medium containing the appropriate antibiotics.
- Incubate at 37 °C and 225 rpm until OD_{600} reaches ~ 0.5 *(depending on the protein to be expressed).*
- Control 1: Pipet 1 ml of the culture into an 1.5 ml tube, centrifuge at 4000 rpm for 5 min, remove the supernatant and resuspend the pellet

in 200 µl of 1x PBS Buffer. Mix with 2x protein sample buffer and store at –20°C.

- Induce protein expression by adding IPTG to a final concentration of 1 mM. Incubate at 30 °C over night.
- Control 2: Measure OD_{600} of the culture. Pipet 1 ml of the culture into a 1.5 ml tube and centrifuge at 4000 rpm for 5 min. Remove the supernatant and suspend the pellet in 200 µl 1x PBS Buffer and lyse the cells, e.g. by sonication. Centrifuge at 18000 rpm for 10 min, take a sample from the supernatant and mix with 2x protein sample buffer. Remove the remainder of the supernatant carefully and resuspend the pellet in the same amount of 1x PBS buffer as before. Take a sample and mix with 2x protein sample buffer. Store the samples at –20 °C.
- For harvesting the cells centrifuge the remainder of the culture for 15 min at 4000 rpm and 4 °C. Remove the supernatant and dissolve the pellet in 20 ml of ice-cold 1x PBS buffer per liter of bacterial culture. *The optimal buffer conditions depend on the properties of the protein of interest and have to be optimized for every purification. E.g. use TrisxHCl buffer for acidic proteins (purified by anion exchange chromatography), or a sodium phosphate buffer for basic proteins (purified by cation exchange chromatography). Lysis buffers can also contain various amounts of salt (10 to 500 mM), reductants as 0.1% (w/v) β-mercaptoethanol or 0.5 to 10 mM DTT, and metal ions if the protein is a metal binding protein.*
- Freeze the bacterial suspension in liquid nitrogen and store at –20 °C over night or –80 °C for long term storage.
- Check the expression of the recombinant protein using SDS-PAGE. Load 10 to 20 µl of the control samples and compare the amount of insoluble and soluble protein fraction.

7.4.3 Preparation of lysates for protein purification

- Thaw an aliquot of the bacterial suspension on ice or in a room tempered water bath.
- Sonicate the lysate on ice.This fragmentizes the bacterial DNA, which makes the lysate rather viscous.
- Centrifuge the lysate at 4 °C and 18000 rpm for 15 min. This should result in a compact pellet and a clear supernatant.
- If not previously done, analyze aliquots of the supernatant and the pellet equivalent to 50 µl of bacterial culture on a SDS Page for the insoluble and soluble protein fraction.

7.4.4 Precipitation of proteins using ammonium sulfate

Analytical method

- For a test precipitation prepare four 200 µl aliquots of lysate containing the protein of interest.

- Add an appropriate amount of solid ammonium sulfate (using a reference table) to the aliquots to adjust to 30%, 40%, 50%, and 60% of ammonium sulfate saturation. Mix and cool on ice for 30 min.
- Centrifuge at 4 °C and 10000 rpm for 30 min. Transfer the supernatants to fresh tubes and supplement to a volume of 500 µl with lysis buffer. Dissolve the pellets in 500 µl of lysis buffer.
- Analyze 5 to 10 µl of these solutions corresponding to 100 to 200 µl of bacterial culture using SDS-PAGE. To remove the ammonium sulfate disturbing electrophoresis precipitate the proteins with trichloric acid as described.
- Determine the ammonium sulfate concentrations where most of the protein of interest remains in the supernatant and where most of the protein is precipitated. If it is necessary test further concentrations.

Fractionated preparative precipitation

- Cool the lysate containing the protein of interest on ice. Add an appropriate amount of solid ammonium sulfate to adjust the concentration where most of the desired protein remains in the supernatant as determined in the analytical experiment. Cool on ice for 30 min.
- Centrifuge at 4 °C and 13000 rpm for 15 min.
- Transfer the supernatant to a fresh tube and again add an appropriate amount of ammonium sulfate to adjust to the concentration where most of the desired protein precipitates. Cool on ice for 30 min.
- Centrifuge at 4 °C and 13000 rpm for 15 min.
- Remove the supernatant and dissolve the protein pellet in a buffer suitable for the next purification step.

7.4.5 Concentration of protein samples

Concentration with ammonium sulfate

Precipitation of proteins under native conditions is a fast and mild method, suitable not only for purification but also for concentration, e.g. after lysis or after chromatographic purification where proteins are sometimes eluted in large volumes.

- Test the ammonium concentration where the protein of interest precipitates quantitatively as described.
- Dissolve the pellet in a buffer suitable for the next purification step or analyses. If the ammonium sulfate remaining in the pellet disturbs the subsequent purification step, remove it by means of size exclusion chromatography using a HiPrepDesalting column (amersham pharmacia biotech).

Concentration using membrane filters

Centriprep or Centricon filters (Amicon) are suitable to generate highly concentrated protein samples, e.g. to obtain protein samples of high concentration for NMR analysis under native conditions.

- Transfer the protein solution to a centriprep (up to 15 ml volume) or a centricon (up to 2 ml volume) with a cut off value smaller than the molecular weight of the protein.
- Centrifuge at 4000 rpm and 4°C, the buffer will pass the filter while the protein will be retained.

7.4.6 Chromatographic Methods

The Fast Performance Liquid Chromatography (FPLC) System

Protein purification was performed using the FPLC system Äkta*explorer* from amersham pharmacia biotech. Chromatographic methods used were ion exchange and size exclusion chromatography. All FPLC purification steps were performed at room temperature.

Size exclusion chromatography for sample desalting

HiPrep Desalting (amersham pharmacia biotech) is a prepacked column for group separation of high (MW > 5000) from low molecular weight substances (MW < 1000), e.g. to separate proteins from salt. The bed volume of the column is 53 ml, the matrix of the column is Sephadex G-25 Fine, cross-linked dextran.

- Choose a buffer system to ensure that the sample is fully soluble. *For substances carrying charged groups a buffer containing up to 150 mM salt (e.g. NaCl) should be used to prevent possible ionic interactions with the matrix. Degas the running buffer using a water jet pump and filtrate through a ZapCapS (Schleicher&Schuell) bottle top filter (0.2 µm).*
- Flush the pumps, the valves and the sample applicator of the FPLC system with the buffer solutions.
- Equilibrate the HiPrep Desalting column with at least 5 CV of running buffer prior to use.
- Dissolve the protein of interest in the running buffer or dialyse to this buffer. Centrifuge at 4°C and 13000 rpm for 10 min or more or filtrate the sample through a filter ring (0.2µm).
- Load the protein solution onto the column by either using a syringe or the sample pump, depending on the sample volume.
- Elute the column with 5 CV of buffer.
- Collect fractions of the flow through and the eluate and analyse the samples on an SDS-PAGE.

- After the separation, flush the FPLC system and wash the column with 5 CV of water and a minimum of 5 CV of 20% ethanol to prevent microbial growth.

Anion exchange chromatography (Resource Q 6ml)

Resource Q is a pre-packed column from pharmacia biotech. The matrix is composed of monodisperse polystyrene/divinyl benzene beads with quaternary ammonium as anion exchanger. It is suitable for purification of acidic proteins (pI< 7). The pH of the buffer should be at least 1 pH unit above the pI of the protein of interest so that it is negatively charged and will bind to the matrix. Anionic detergents should be avoided as they bind to quaternary ammonium. The practical protein loading range is up to 150 mg, whereas the sample volume is of minor importance.

- Choose a buffer system that is not bound by the exchanger (e.g.: Tris×HCl) and prepare two buffer solutions, one with low ionic strength (e.g. 10 mM NaCl) and one with high ionic strength (e.g. 0.5 to 1 M NaCl).
- Degas the buffer solutions using a water jet pump and filtrate through a ZapCapS bottle top filter (0.2 µm).
- Flush the pumps, the valves and the sample applicator of the FPLC system with the buffer solutions.
- Equilibrate the column with at least 2 CV of start buffer (low ionic strength buffer).
- Run 2 CV of elution buffer (high ionic strength buffer) through the column.
- Re-equilibrate the column with 5 CV of start buffer.
- Centrifuge the bacterial lysate at 4°C and 18000 rpm for at least 10 min to remove insoluble particles.
- Load the supernatant onto the column using the sample pump.
- Wash with 5 CV of start buffer.
- Elute the protein via gradient elution by mixing start and elution buffer over e.g. 20 CV lower flow rates and more shallow gradients improve resolution.
- Collect fractions of the flow through and the eluate and analyse the samples by SDS-PAGE.
- After the purification, flush the FPLC system, wash the column with 5-10 CV of H_2O and eventually with ~ 5 CV of 20% ethanol for storage. From time to time clean the column with 5 CV 1 M NaCl, 5 CV 1 M NaOH, 5 CV 1 M HCl and 5 CV 1 M NaCl, before washing with H_2O and storing in 20% ethanol.

Size exclusion chromatography using HiLoad Superdex columns

HiLoad 16/60 Superdex 75 prep grade is a prepacked column from amersham pharmacia biotech. The matrix is produced by covalent

bonding of dextran to highly cross-linked agarose. The separation range of this column lies in the molecular weight range between $3x10^3$ to $7x10^4$, the bed volume is approximately 120 ml. Proteins applied to this column are separated according to their size. Large proteins elute first, while small proteins can enter the pores of the matrix are thus retarded. This polishing step is mostly used as a last step in protein purification to achieve final high-level purity.

- Choose a buffer system that is suitable for the analyses following the purification (e.g. sodium or potassium phosphate at appropriate pH). Degas the buffer solutions using a water jet pump and filtrate through a ZapCapS bottle top filter (0.2 µm).
- Equilibrate the column with at least 2 CV of running buffer until the base line is stable.
- Flush the pumps, the valves and the sample applicator of the FPLC system with the buffer solutions.
- Dissolve the protein sample in 1 to 2 % of the bed volume (1 to 1.5 ml) of running buffer or concentrate to this volume.
- Centrifuge the protein sample at 4 °C and 18000 rpm for at least 10 min.
- Load the protein solution onto the column using a syringe.
- Elute the column with 1 bed volume of running buffer.
- Collect fractions of the flow through and the eluate and analyse the samples on an SDS-PAGE.

If the resolution is low, decrease the flow rate or the sample volume or amount. Proteins with polar residues on their surface may interact with the matrix and are poorly eluted. In this case use buffers with higher ionic strength, containing e.g. 150 mM NaCl. Alternatively, gel filtration can be performed under denaturing conditions with buffers containing up to 8 M urea or up to 6 M guanidinium hydrochloride.

- After the purification, flush the FPLC system and wash the column with a minimum of 3 CV of water and eventually with 4 CV of 20% of ethanol.

7.4.7 Quantification of protein

Coomassie Blue protein assay (Bradford assay)

- The absorbance maximum of the dye in an acidic solution shifts from 465 to 595 nm after adding protein due to stabilization of the anionic form of the dye by both hydrophobic and ionic interactions. The dye principally reacts with arginine residues and to a lesser extent with histidine, lysine, tyrosine, and phenylalanine residues.
- Place 20 µl of the samples I n a 1.5 ml tube.
- Add 1 ml of Bradford reagent.

- Immediately measure the absorbance at 595 nm in polystyrene cuvettes.
- For calibration prepare several dilutions of BSA with protein concentrations between 10 and 1000 µg/ml. Treat them in the same way as the samples.

UV spectrometry: absorbance at 280 nm

Quantitative study of protein-protein and protein-DNA interactions in solution requires accurate determination of protein concentration. In contrast to the Bradford assay, measurement of the absorbance at 280 nm results in a quite accurate determination of the protein concentration. If the extinction coefficient has not yet been determined experimentally, it is possible to calculate the molar extinction coefficient simply from knowledge of the amino acid composition.[3] Following amino acid residues contribute to the extinction coefficient of a protein: Trp (5690 $M^{-1}cm^{-1}$), Tyr (1280 $M^{-1}cm^{-1}$), Cys (120 $M^{-1}cm^{-1}$).

- Search for Trp, Tyr, and Cys residues in the amino acid sequence of the protein. Multiply the numbers of these amino acid residues in the protein with their contributions to the extinction coefficient of the protein.
- Measure the absorbance of the denatured protein in 6 M guanidiumhydrochloride at 280 nm. Choose the concentration of the protein so that the absorbance is about 0.5.
- Calculate the concentration of the protein according to the Lambert-Beer law: $A = \varepsilon.c.d$, with A being the measured absorbance, ε the molar extinction coefficient of the protein, c the concentration of protein, and d the path length.

7.4.8 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) of proteins

SDS-PAGE is used for investigation of protein samples in terms of contained proteins. Gels containing sodium dodecyl sulfate (SDS) provide dissociating conditions, where proteins are separated strictly according to polypeptide size. A discontinuous buffer system is used.[5, 6] Polyacrylamide (PAA) concentration is 15% for all gels, the molecular weight marker (Dalton Mark VII-L Standard Mixture, Sigma Aldrich) range is from 14 to 66 kDa showing bands at 14.2, 20.1, 24.0, 29.0, 36.0, 45.0, and 66.0 kDa. Gels are cast using the Mini-PROTEAN 3 Multi-Casting Chamber (BIORAD).

	12 resolving gels	12 stacking gels

Tris (pH 9.0)	8 ml	-
Tris (pH 6.8)	-	3.5 m
40% PAA (37.5 : 1)	24 ml	-
40% PAA (19 : 1)	-	3.5 m
H_2O	32ml	21ml
	Degas	
20% SDS	320 µl	70 µl
TEMED	20 µl	35 µl
APS	200 µ	140 µl

- Put the gel into a running cell filled with Laemmli buffer.
- Load the sample onto the gel.
- Run the gel at 200 V for 40 min.
- Incubate the gel in Coomassie staining solution for at least 30 min.
- Destain the gel in destaining solution until the gel is colourless again, except the protein bands.
- Take a picture of the gel with a Gel Doc System (Sony).

7.4.9 Circular dichroism spectroscopy of proteins

Circular dichroism (CD) is defined as the difference in absorption of left and right handed circular-polarized light $\Delta A = A_L - A_R$ by chromophores in an asymmetrical environment, resulting in elliptically polarized light. CD spectroscopy can be used to get information about the structure of macromolecules (including the secondary structure of proteins and the handedness of DNA). The most commonly used unit in CD spectroscopy of proteins is the mean residue ellipticity [kdeg cm² dmol^{-1}], which is obtained by normalization with sample concentration, path length of the cell and with the number of residues. The mean residue ellipticity is measured as a function of wavelength.

For gaining information about the secondary structure of proteins, CD spectra need to be recorded from about 260 nm to ~ 180 nm (far UV). At these wavelengths the amide groups of the peptide backbone interact with the entering circularly-polarized light. Spectra can be analyzed for different secondary structural types: alpha helix, parallel and antiparallel beta sheet, random coil, etc. shows typical CD spectra of different secondary structural elements.

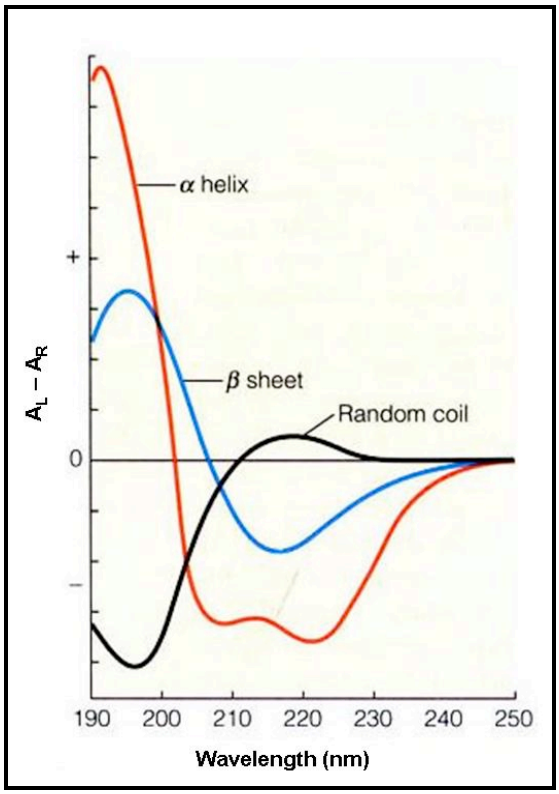


Fig. 1. Typical CD spectra of different secondary structural elements (far UV). Spectrum α-helix (red); β-sheet (blue); random coil (black). The ordinate shows the difference in absorption of left and right handed circular-polarized light $\Delta A = A_L - A_R$ as a function of the wavelength in nm.

Experimental conditions:

Protein concentrations were determined by photometry at 280 nm. Samples containing 0,5 mM protein dissolved in a sodium phosphate buffer (100 mM NaH2PO4/Na2HPO4, 150 mM NaCl, adjusted to pH 6.5) were analysed with an “Applied Photosystems Π*-180” spectropolarimeter using quartz cells of 1 mm path-length. Spectra were processed by subtracting buffer spectra measured under the same conditions. Wavelength spectra were obtained at 20°C recorded at 0.5 nm intervals from 180 to 260 nm. Using the base spectra of 33 reference proteins, quantitative determination of secondary structure elements of the protein in question was achieved.

Temperature scans were performed from 20 to 60°C and the ellipticity at 218 nm was recorded.

7.5 Specific Protocoll: Osteopontin Expression and Purification

Bacteria of the E. Coli strain BL21(DE3) transformed with pET11d-OPN220 expression vector were grown in 20ml LB until an OD_{600} of 0.5 was reached. The culture was then added to 1l M9 media supplemented with Ampicilline. The bacteria were grown to an OD_{600} of 0.5 when recombinant protein synthesis was induced with IPTG to a final concentration of 1mM. The bacteria were grown overnight at 30°C and are

ready to harvest on the next day having reached an OD_{600} of 1.5-2. The cells were collected by centrifugation at 4000rpm for 20min and resuspended in 20ml of PBS-Buffer supplemented with 1ml Protease Inhibitor Cocktail (Sigma). Bacteria were lysed using a sonicator (15min) and the cell lysate was cleared by centrifugation at 18000rpm for 20min. The supernatant containing the soluble protein fraction was adjusted to a $(NH_4)_2SO_4$ concentration of 40% and cleared by centrifugation (18000rpm/20min). The soluble fraction was again adjusted to 50% $(NH_4)_2SO_4$. After another round of centrifugation the precipitate containing most of the OPN220 protein was resuspended in 10ml PBS buffer and purified form residual $(NH_4)_2SO_4$ with a desalting column. The fraction containing protein were pooled and loaded onto a FPLC anion exchange column (ResourceQ). Bound protein was eluted by gradient mixing PBS buffer with PBS buffer containing 1M NaCl. The fractions containing OPN220 were pooled and concentrated to 1ml with an Amicon centrifugal filter device. (10kDa NMWL) and then loaded onto a Superdex Hiload 16/60 gel filtration column (GE Healthcare) equilibrated in NaP Buffer. For NMR analysis, protein samples were concentrated to 0.5mM.

This protocol also applies to single cystein mutant OPN220 purification. Only 1mM of DTT was added to all buffers to avoid the formation of internuclear disulfide bonds.

All purification steps should be executed as quickly as possible since OPN is degrading very fast in solution. Storage overnight during the purification process should be at -20°C.

7.6 Protein NMR

7.6.1 Introduction

The technique of NMR takes advantage of a specific property of certain nuclei, the nuclear spin. The magnetic resonance phenomenon occurs as a result of the quantum mechanical property of the spin, where the resulting angular momentum confers a magnetic moment on the nucleus. Introduction into a static magnetic field leads to a splitting up into different energy levels amenable for the magnetic moment. Irradiation with radio-frequency pulses leads to excitation of these magnetic moments into levels of higher energy and generates detectable signals.

7.6.2 NMR experiments

All NMR diffusion measurements were performed on a Varian Inova spectrometer, operating at 600 MHZ.

NMR experiments are used to obtain physical, chemical and structural information of a biological macromolecule. These experiments can provide detailed information on the topology, dynamics and three-dimensional structure of molecules in solution (and the solid state). NMR Experiments are built up of distinct building blocks. A short summary of some basic building blocks for NMR experiments used in this work and their principles is given in the following section.

NMR building blocks

Insensitive Nuclei Enhanced by Polarization Transfer (INEPT)

The INEPT[7] sequence transfers *I* spin polarization from a sensitive nucleus with a high gyromagnetic ratio, γI (usually protons, ^{1}H) to *S* spin polarization of a less sensitive nucleus with a lower gyromagnetic ratio, γS (e.g. nitrogen, ^{15}N) by means of scalar coupling interaction. The experiment yields anti-phase S spin magnetization with respect to spin *I*. The INEPT sequence is the most frequently used method for heteronuclear magnetization transfer in NMR experiments in solution with indirect detection of the heteronucleus and increases the sensitivity of the experiment by $\gamma H/\gamma N \approx 10$.

Sensitivity Enhancement (SE)

Sensitivity enhancement[13] employs the Cavanagh-Rance trick[14] to convert both coherence transfer pathways which can be obtained in the indirect dimension into observable magnetization. Instead of alternatively acquiring the cosine and the sine-modulated signal in the indirect dimension, both types of modulations are observed simultaneously by refocusing the unobservable modulation in an additional spin-echo sequence. SE can provide sensitivity improvements by factors up to $\sqrt{2}$ relative to conventional experiments. However, due to relaxation during the extra delays in the experiment and due to radio-frequency inhomogenity in the additional pulses, this theoretical sensitivity enhancement is usually not obtained in practice.

Decoupling

Coupling (scalar and dipolar) between two spins results in peak splitting, thus leading to overcrowded spectra with reduced peak intensities. In order to circumvent this problem, the coupling partner has to be decoupled during the evolution and acquisition period. This is achieved by applying a 180° pulse in the center of the evolution period and a series of 180° pulses during the acquisition of the FID. The observed peak resembles the average frequency of the split peaks.

Heteronuclear NMR experiments

2D ^{1}H-^{15}N HSQC

This experiment allows the correlation of amide protons with their directly attached nitrogen. An INEPT pulse sequence transfers proton magnetization (H_z) into nitrogen antiphase heteronuclear single-quantum coherence ($2H_zN_y$). ^{15}N frequency labeling during a t_1 evolution period is followed by a refocused reverse INEPT transfer, and subsequent sensitivity-enhancement generates detectable proton magnetization. The magnetization pathway is as follows:

$^{1}H^{N} \rightarrow \ ^{15}N\ (t_1) \ \rightarrow \ ^{1}H^{N}\ (t_2)$

2D ^{15}N T_1 experiment

The ^{15}N T_1 experiment[15] is basically an HSQC experiment with an additional variable delay for specific ^{15}N longitudinal relaxation. The signal intensities scale with the loss of magnetization due to longitudinal relaxation during this delay. Increasing its duration in a series of experiments, relaxation decay curves can be used to determine ^{15}N T_1 times. After INEPT transfer and subsequent refocusing, longitudinal ^{15}N magnetization is generated. Relaxation of ^{15}N occurs via dipolar coupling between ^{1}H and ^{15}N during a variable period. Effects of cross-correlation relaxation mechanisms between ^{1}H - ^{15}N dipolar and ^{15}N CSA coupling are eliminated through ^{1}H 180° pulses during the relaxation time. At the end of this period longitudinal ^{15}N magnetization is transferred back into transverse ^{15}N magnetization, and the coherence pathway proceeds as described in the HSQC experiment. The magnetization pathway is as follows:

$^{1}H^{N} \rightarrow \ ^{15}N \xrightarrow{R_1} \ ^{15}N\ (t_1) \ \rightarrow \ ^{1}H^{N}\ (t_2)$

2D ^{15}N T_2 Experiment

The ^{15}N T_2 experiment[15] is basically an HSQC experiment with an additional variable delay (cycles) for specific ^{15}N transverse relaxation. The signal intensities scale with the loss of magnetization due to transverse relaxation during this delay. Increasing its duration in a series of experiments, relaxation decay curves can be used to determine ^{15}N T_2 times. After INEPT transfer and subsequent refocusing transverse ^{15}N magnetization is generated. Transverse relaxation occurs during the application of a Carr-Purcell-Meiboom-Gill (CPMG) spin-echo sequence[16, 17] with successive 180° ^{15}N pulses to refocus transverse magnetization. Effects of cross-correlation relaxation mechanisms between ^{1}H-^{15}N dipolar and ^{15}N CSA coupling are eliminated through ^{1}H 180° pulses during the relaxation time.[18] At the end of this period transverse ^{15}N magnetization

proceeds as described in the HSQC experiment. The magnetization pathway is as follows:

$^{1}H^{N} \rightarrow {}^{15}N \xrightarrow{R_2} {}^{15}N\,(t_1) \rightarrow {}^{1}H^{N}\,(t_2)$

CPMG relaxation dispersion experiment

Many biological events occur on the milli- to microsecond timescale. The Carr−Purcell−Meiboom−Gill (CPMG) technique method can be used to characterize conformational transitions that are involved in such processes.

The CPMG relaxation dispersion experiment applies a variable number of refocusing pulses during a fixed time interval. If the rate of application of pulses is such that the magnetization is allowed to get out of phase, then the exchange process leads to broad lines. In contrast, when pulses are applied at a fast rate, the exchange process is quenched and narrower peaks are obtained. The dependence of the peak height on the pulsing rate is subsequently used to extract exchange information about the protein in question.

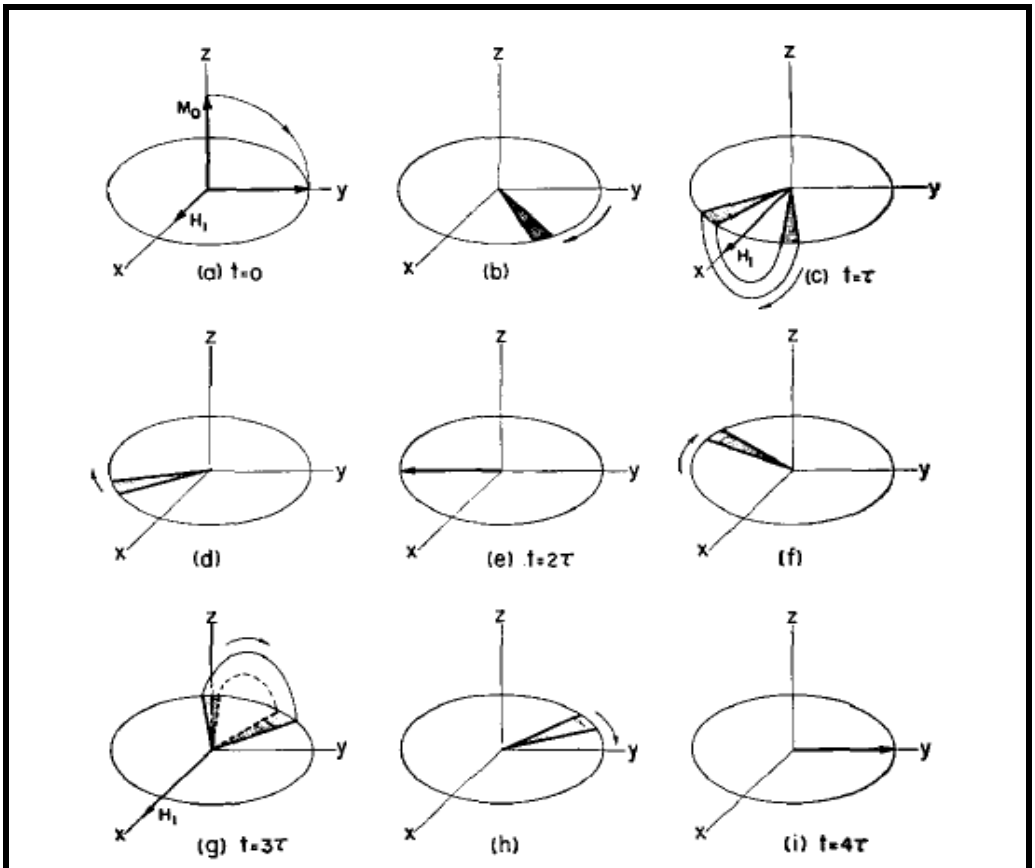


FIG. 1. Behavior of the nuclear polarization in a spin-echo scheme with coherent pulses. A 90° pulse is applied at time $t=0$ and 180° pulses at times $t=(2n+1)\tau$, $(n=0, 1, 2 \cdots)$.

IPAP experiment for measuring one-bond ^{15}N–$^{1}H^{N}$ residual dipolar coupling (RDC)

RDCs arise when molecular systems containing proximate pairs of magnetic nuclei are partially ordered in magnetic fields. The underlying mechanism is the same as the through-space dipole-dipole coupling that dominates solids NMR spectra. For a pair of spin 1/2 nuclei in a magnetic field, the distance and angle dependence are shown in the equation below, where r is the distance between a specific pair of nuclei, $\gamma_{i,j}$ are the gyromagnetic ratios for the nuclei, μ_0 is the permittivity of space, h is Planck's constant, and θ is the angle between the considered internuclear vector and the magnetic field.

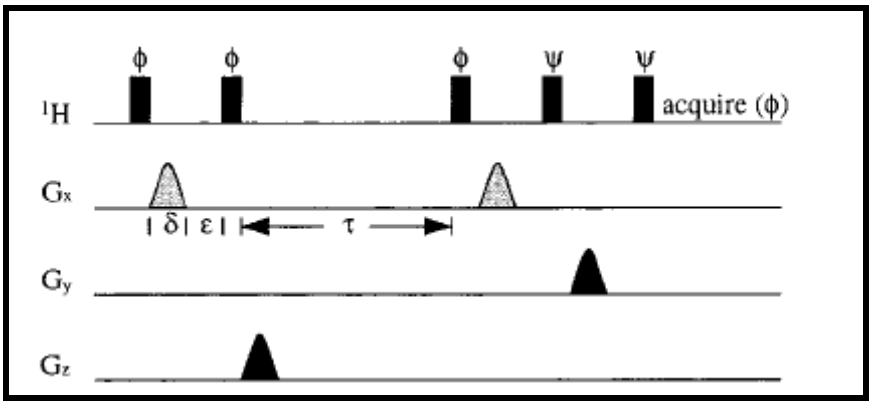
$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h}{(2\pi r)^3}\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle$$

When all parameters are given in SI units, the resulting D_{ij} is given in Hertz. Many measurements of RDCs are made between pairs of bonded nuclei, so that r is fixed; RDCs have, thus, been used primarily to provide angular information. Because of the identical spin operators, RDCs add to scalar couplings (J_{HNN}) to produce splittings of $J_{HNN} + D_{HNN}$.

To obtain ^{15}N–$^{1}H^{N}$ couplings, a ^{15}N HSQC without decoupling the proton during nitrogen evolution is recorded. The resulting signal splitting leads to a doubling of peaks. This might result in a crowded spectrum with signal overlap. To circumvent this problem, two separate spectra can be acquired. One contains the **i**n**p**hase component of the coupled magnetization, on the **a**nti**p**hase component. (**IPAP**) Both spectra are added and subtracted from each other to yield the upfield and the downfield component of the doublet. Splittings are measured under isotropic (J_{HNN}) and aligned ($J_{HNN} + D_{HNN}$) conditions to isolate the RDC contribution

PFG NMR Hydrodynamic Radius Measurements

In the characterization of native proteins in solution, a measurement of the molecular dimensions of the system is of considerable importance. PFG (Pulsed Field Gradient) NMR methods are designed to measure the hydrodynamic radii of proteins. Dioxan was added to the protein solutions as an internal standard. PFG NMR diffusion measurements were performed with the PG-SLED (pulse gradient stimulated echo longitudinal encode-decode) sequence:



The lengths of all pulses and delays in this sequence were held constant and 50 spectra were acquired with the strength of the diffusion gradient varying between 5% and 100% of its maximum value. The lengths of the diffusion gradient and the stimulated echo were optimized for each sample to give a total decay in the protein signal of between 80% and 90%.

The diffusion gradient time (δ) used was 4.5 ms and the echo time (Δ) was 100 ms. 50 experiments were recorded with increasing diffusion gradient strength. The data were processed using NMRPipe[26]. The decay of the protein aliphatic proton signal was fitted to a single Gaussian using the DOSY module of NMRPipe in order 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_{ref}). The hydrodynamic radius of the protein (R_h^{prot}) was deduced from:

$$R_h^{prot} = \frac{d_{ref}}{d_{prot}}(R_h^{ref})$$

assuming R_h^{diox} (hydrodynamic radius of dioxane) is 2.12 Å [57].

Small-angle x-ray scattering.

The SAXS equipment consisted of a slit-geometry camera with high flux and low background (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 (λ = 0.154 nm). A CCD camera, from Princeton instruments, a division of Roper Scientifics (Trenton, NJ, USA), recorded the 2D scattering pattern. The temperature of the capillary in the metallic sample holder was controlled by a Peltier element. Measurements were carried out at 25°C. The measuring times were 3×10 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 $q = (4p/l)\sin(q/2)$, where q 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 made using water as a secondary standard.

The obtained SAXS spectra were further analyzed with the indirect Fourier transformation method IFT. 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 the real space:

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

Here, 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, 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.

7.6.3 Identifying secondary structure elements

The chemical shift

In magnetic resonance spectroscopy the electrons in the molecules cause the local magnetic field to vary on a submolecular distance scale. Thus, the magnetic fields experienced by nuclei at two sites in the same molecule are different if the electronic environments are different. This effect is called the chemical shift. It is very useful as it allows the atomic nuclei to send out a variety of molecular information encoded as radio waves. It is highly sensitive to many atomic or molecular effects including local electron distributions, bond hybridization states, proximity to polar groups, nearby aromatic rings and local magnetic anisotropies. All those effects may, in turn, be dependent on the atomic composition, the molecular geometry or the solvent constitution. The origin of chemical shift dispersion is anisotropic diamagnetic susceptibilities of molecular fragments in the polypeptide chain:

- Backbone chemical shifts are influenced by the backbone dihedral angle ϕ and, although less pronounced, by ψ, as well as by anisotropy effects such as those of carbonyl double bonds.
- Particularly large effects are due to the local ring current fields of nearby aromatic systems. Ring current shifts arise because spins above or below an aromatic ring plane experience a local field (B_{loc}) that opposes the static magnetic field B_0 and therefore experience an upfield shift. Conversely, spins located in the ring plane experience a downfield shift.
- Amide chemical shifts are sensitive to hydrogen bonding. In detail, it has been shown that backbone amide H^N chemical shifts are related to the hydrogen bond length and strength, wherein short (i.e. strong) bonds lead to downfield shifts and long (i.e. weak) bonds lead to upfield shifts.

Random coil chemical shifts

Random coil chemical shifts, $\delta_{rc,}$ represent the chemical shift of individual residues within completely flexible, unconstrained peptides. They fall roughly halfway between the chemical shift extremes found for residues in α-helices and β-strands. Wishart and coworkers presented ^{1}H, ^{15}N and ^{13}C random coil chemical shift values of the common amino acids. Hexa-peptides of the form gly-gly-x-ala-gly-gly were used as model systems, where x represents one of the 20 naturally occurring amino acids. This study provides a complete and internally consistent set of ^{1}H, ^{15}N, and $^{13}C\alpha$ random coil chemical shift values.

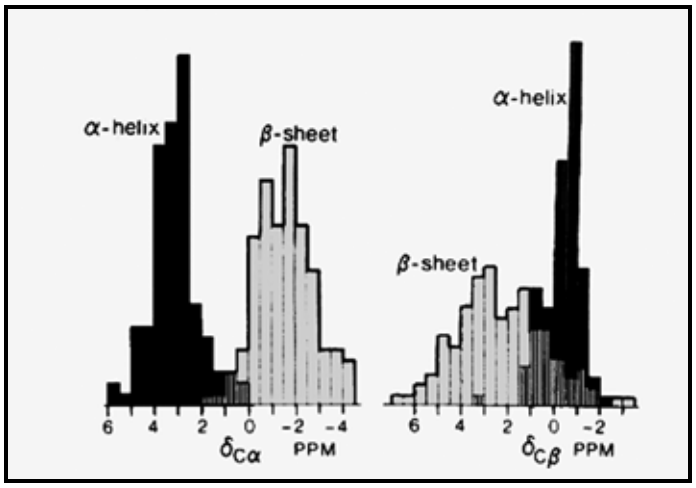
Secondary chemical shifts

Since the 1960s there were a number of attempts on understanding the conformational and constitutive effects on chemical shifts of amino acids in polymer systems. These efforts were made in an attempt to determine the three-dimensional structure of proteins through the precise understanding of the geometric dependencies of chemical shifts. First simple theories concerned the prediction of shift tendencies in α-protons upon helix-formation and further attempts were done until the early 1980s. In the 1990s a number of structures were already solved using newly developed multidimensional NMR methods. This provided a flood of structural data, which was amenable to computational analysis. The results of these studies revealed a clear structural relationship for all 20 amino acids for not only α-protons but for amid protons, α-carbon atoms, carbonyl atoms and amide nitrogen atoms.

In general, observed chemical shifts in proteins can be partitioned into the sum of two components: the random coil chemical shifts, $\delta_{rc,}$ and the conformation-dependent ‘secondary chemical shifts’, $\Delta\delta$.

$$\delta_{obs} = \delta_{rc} + \Delta\delta$$

Unlike the random coil chemical shift, the secondary chemical shift contains the contribution from secondary and tertiary structure. Especially, $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts were found to be predominantly determined by the backbone ϕ,ψ dihedral angles.



The average deviation from random coil chemical shifts of $^{13}C\alpha$ is 3.09 ± 1.00 ppm for residues located in α-helices and –1.48 ± 1.23 ppm for residues in β-strands, representing a significant downfield shift for helical residues and an upfield shift for residues located in β-strands. Conversely, the average $^{13}C\beta$ shift deviates –0.38 ± 0.85 ppm for α-helical and 2.16 ± 1.91 ppm for β-strand conformation.

^{13}CO secondary shifts are more influenced by residue type and by the local amino acid sequence and remain well dispersed, even in the unfolded state. Their behavior is similar to $^{13}C\alpha$ secondary shifts with a downfield shift for residues in helical structures and upfield shifts for residues in β-strands. The situation for amide ^{15}N is less clear. Nitrogen chemical shifts are found to display very broad chemical shift dispersion and despite the average shift difference between α-helices and β-strands being more than 3 ppm, the overlap between the two conformational groups is often larger than 3 ppm. Due to this the ^{15}N nucleus is not expected to be as informative as, for example, carbon.

7.6.4 Relaxation mechanisms and molecular motions

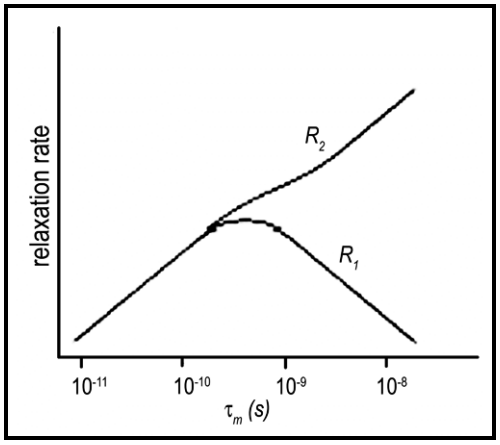
NMR relaxation studies can provide detailed information pertaining to the internal dynamics occurring in proteins. The measurement of ^{15}N or ^{13}C relaxation rates is particularly useful to obtain dynamic information since the relaxation of these nuclei is governed predominantly by the dipolar interaction with directly bound protons and to a much smaller extent by the chemical shift anisotropy mechanism. Relaxation in molecules principally occurs either according to the longitudinal or spin-lattice relaxation

process or according to the transverse or spin-spin relaxation process. Both lead to a decay of the magnetization in an excited spin system and constitute the phenomenon of free induction decay, where the time constant of the particular contribution is the reciprocal of the relaxation rate, *R*. However, the two mechanisms are distinctly different.

The longitudinal relaxation is an enthalpic process and is mediated by energy exchange with the 'lattice', i.e. the sum of translational, rotational and internal motions of the molecule. In order to enable the transition form the excited state to the ground state, the 'lattice' must cause a fluctuating field, whose frequency corresponds to the transition. The fluctuating field itself is determined by the specific dynamic parameters of the protein, such as overall rotational correlation time, internal correlation time and chemical exchange processes. The time dependent orientation of a N-H bond vector in a protein can be described by a correlation function of the form

$$C(t) = e^{-t/\tau_m}$$

where the correlation time τ_m represents the time required for the molecule to rotate by roughly one radian about any axis, assumed for isotropic molecular tumbling (no preferred axis of rotation). Fourier transformation of this correlation function results in the spectral density function, $J(\omega)$, which gives information about the power available from molecular motions to bring about relaxation as a function of the molecular tumbling rate. For the R_1 mechanism, the relaxation maximum is at $\omega\tau_m \approx 1$, where ω is the resonance frequency in radian units for the transition of the investigated nucleus.



Rate constants R_1 and R_2 as a function of the overall tumbling correlation time τ_m.

The transverse relaxation is an entropic process and is mediated by mutual exchange of spin energies (mutual spin flip), with no net change in energy. This mechanism predominantly occurs due to fluctuating magnetic fields when $\omega\tau_m \gg 1$.

The heteronuclear $\{^1H\}$-^{15}N-NOE arises due to energy level redistribution of dipole-dipole coupled spins upon perturbation of the equilibrium state. Double- and zero- quantum transitions lead to a decrease or increase in the population difference of a particular spin, and an enhancement or decrease of the corresponding NMR signal arises. Thereby, double-quantum (high frequency) transitions lead to a positive NOE enhancement, whereas zero-quantum (low frequency) transitions lead to a negative enhancement. Consequently, the NOE is dependent on the present fluctuating magnetic fields and thus dependent on the correlation time τ_m. Motional behavior of the N-H bond vector can be observed directly by the sign and the magnitude of the $\{^1H\}$-^{15}N-NOE. Nevertheless, NOE enhancement is determined by comparing signal intensities of the individual resonances with and without precedent water saturation and thus requires nicely separated signals in order to exclude mutual interference from overlapping signals.

^{15}N T_1 and T_2 relaxation times are determined from ^{15}N T_1 and ^{15}N T_2 NMR experiments. The delay, during which specific longitudinal and transverse relaxation occur, respectively, is varied in a series of experiments. The signal intensities in each spectrum, I, scale with the loss of magnetization during this delay, t, and relaxation curves for each residue are then obtained by plotting the signal intensity versus the duration of the delay. The following equation is used in a curve fit program to extract T_1 and T_2:

$$I(t) = A\, e^{-t/T_{1,2}}$$

where the two variables are the scaling factor A and the relaxation time T_1 and T_2. The heteronuclear $\{^1H\}$-^{15}N-NOE is determined as the ratio of the crosspeak intensities in two experiments, one with and one without presaturation of the amide proton resonances.

$$NOE = I_{sat}/I_{nosat}$$

8 References

[1] Fedarko, Jain, Karadag, Fisher (2004). Three small integrin binding ligand N-linked glycoproteins (SIBLINGs) bind and activate specific matrix metalloproteinases. FASEB Journal, **18**, 734–736

[2] Fisher, Whitson, Avioli, Termine (1983). Matrix sialoprotein of developing bone. J. Biol. Chem. **258**, 12723–12727

[3] George, Sabsay, Simonian, Veis (1993). Characterization of a novel dentin matrix acidic phosphoprotein. Implications for induction of biomineralization. J. Biol. Chem. **268**, 12624–12630

[4] Hunter, Hauschka, Poole, Rosenberg, Goldberg (1996). Nucleation and inhibition of hydroxyapatite formation by mineralized tissue proteins. Biochem. J. **317**, 59–64

[5] Bellahcene, Castronovo (1997). Expression of bone matrix proteins in human breast cancer: potential roles in microcalcification formation and in the genesis of bone metastases. Bull. Cancer 84, 17–24

[6] Bell (1996). The fibrinolytic system in neoplasia. Semin. Thromb. Hemost **22**, 459–478

[7] Price, Thompson (2002). Mechanisms of tumor invasion and metastasis: emerging targets for therapy. Expert Opin. Ther. Targets. **6**, 217–233

[8] Mi, Guo, Wai, Gao, Kuo (2006). Integrin-linked kinase regulates osteopontindependentMMP-2 and uPA expression to conveymetastatic function in murine mammary epithelialcancer cells. Carcinogenesis **27**, 1134–1145

[9] Das, Mahabeleshwar, Kundu (2004). Osteopontin induces AP-1-mediated secretion of urokinase-type plasminogen activator through c-Srcdependent epidermal growth factor receptor transactivation in breast cancer cells. J. Biol. Chem. **279**, 11051–11064

[10] Teti et al (1998). Activation of MMP-2 by human GCT23 giant cell tumour cells induced by osteopontin, bone sialoprotein and GRGDSP peptides is RGD and cell shape change dependent. Int. J. Cancer **77**, 82–93

[11] Rangaswami, Bulbule, Kundu (2006). Nuclear factor inducing kinase: a key regulator in osteopontininduced MAPK/IκB kinase dependent NF-κB-mediated promatrix metalloproteinase-9 activation. Glycoconj. J. **23**, 221–232

[12] Fedarko, Jain, Karadag, Fisher (2004). Three small integrin binding ligand N-linked glycoproteins (SIBLINGs) bind and activate specific matrix metalloproteinases. FASEB J. **18**, 734–736

[13] Takafuji, Forgues, Unsworth, Goldsmith, Wang (2007). An osteopontin fragment is essential for tumor cell invasion in hepatocellular carcinoma. Oncogene **26**, 6361–6371.

[14] Fidler (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nature Rev. Cancer **3**, 453–458

[15] Oates, Barraclough, Rudland (1996). The identific ation of osteopontin as a metastasis-related gene product in a rodent mammary tumor model. Oncogene **13**, 97–104

[16] Bellahcene, Merville, Castronovo (1994). Expression of bone sialoprotein, a bone matrix protein, in human breast cancer. Cancer Res. **54**, 2823–2826

[17] Waltregny et al (1998). Prognostic value of bone sialoprotein expression in clinically localized human prostate cancer. J. Natl Cancer Inst. **90**, 1000–1008

[18] Bellahcene et al (1997). Expression of bone sialoprotein in human lung cancer. Calcif. Tissue Int. **61**, 183–188

[19] Bellahcene et al (2000). Bone sialoprotein mRNA and protein expression in human multiple myeloma cell lines and patients. Br. J. Haematol. **111**, 1118–1121

[20] McKee & Nanci (1996). Secretion of Osteopontin by macrophages and its accumulation at tissue surfaces during wound healing in mineralized tissues: a potential requirement for macrophage adhesion and phagocytosis. *Anat. Rec.* **245**, 394–409

[21] O'Regan *et al* (1999). Osteopontin is associated with T cells in sarcoid granulomas and has T cell adhesive and cytokine-like properties *in vitro*. *J. Immunol.* **162**, 1024–1031

[22] Weber & Cantor (1996). The immunology of η-1/ osteopontin. *Cytokine Growth Factor Rev.* **7**, 241–248

[23] Giachelli, Lombardi, Johnson, Murry, Almeida (1998). Evidence for a role of osteopontin in macrophage infiltration in response to pathological stimuli *in vivo*. *Am. J. Pathol.* **152**, 353–358

[24] Kawamura *et al* (2005). Differentiation, maturation, and survival of dendritic cells by osteopontin regulation. *Clin. Diagn. Lab. Immunol.* **12**, 206–212

[25] Koh *et al* (2007). Role of osteopontin in neutrophil function. *Immunology* **122**, 466–475

[26] Volanakis (1998). The Human Complement System in Health and Disease 9–32 (Marcel Dekker, New York)

[27] Fedarko, Fohr, Robey, Young, Fisher (2000). Factor H binding to bone sialoprotein and osteopontin enables tumor cell evasion of complement-mediated attack. *J. Biol. Chem.* **275**, 16666–16672

[28] Jain, Karadag, Fohr, Fisher, Fedarko (2002). Three SIBLINGs (small integrin-binding ligand, N-linked glycoproteins) enhance factor activity enabling MCP-like cellular evasion of complement-mediated attack. *J. Biol. Chem.* **277**, 13700–13708

[29] Nam *et al* (2006). Bone sialoprotein mediates thetumor cell-targeted prometastatic activity of transforming growth factor β in a mouse model of breast cancer. *Cancer.* **66**, 6327–6335

[30] Brooks, Clark, Cheresh (1994). Requirement of vascular integrin αvβ3 for angiogenesis. *Science* **264**, 569–571

[31] Cui *et al* (2007). Abrogation of the interaction between osteopontin and αvβ3 integrin reduces tumor growth of human lung cancer cells in mice. *Lung Cancer* **57**, 302–310

[32] Tang *et al* (2007). Inhibition of osteopontin would suppress angiogenesis in gastric cancer. *Biochem. Cell Biol.* **85**, 103–110

[33] Rice, Courter, Giachelli, Scatena (2006). Molecular mediators of αvβ3-induced endothelial cell survival. *J. Vasc Res.* **43**, 422–436

[34] Chakraborty, Jain, Kundu (2008). Osteopontin promotes vascular endothelial growth factordependentbreast tumor growth and angiogenesis via autocrine and paracrine mechanisms. *Cancer Res.* **68**, 152–161

[35] Giachelli (2005). Inducers and inhibitors of biomineralization: lessons from pathological calcification. *Orthod. Craniofac. Res.* **8**, 229–231.

[36] Carlinfante *et al (2005).* Differential expression of osteopontin and bone sialoprotein in bone metastasis of breast and prostate carcinoma. *Clin. Exp. Metastasis* **20**, 437–444

[37] Tunio Hirota, Nomura, Kitamura (1998). Possible relation of osteopontin to development of psammoma bodies in human papillary thyroid cancer. *Arch. Pathol. Lab. Med.* **122**, 1087–1090

[38] Gericke, Qin, Spevak, Fujimoto, Butler, Sorensen, Boskey (2005). Importance of phosphorylation for osteopontin regulation of biomineralization. Calcif Tissue Int. **77**, 45-54

[39] Gokhale, Robey, Boskey (2001) The Biochemistry of bone. Osteoporosis, Second Edition vol 1, Academic Press, San Diego. 107–189

[40] Addadi, Weiner, Geva (2001). On how proteins, interact with crystals and their effect on crystal formation. Z Kardiol. **90**, 92–98.

[41] Furger *et al (2003)* β3 integrin expression increases breast carcinoma cell responsiveness to the malignancy-enhancing effects of osteopontin. *Mol. Cancer Res.* **1**, 810–819.

[42] Rangaswami, Bulbule, Kundu (2006). Osteopontin: role in cell signaling and cancer progression. *Trends Cell Biol.* **16**, 79–87.

[43] Teramoto *et al (2005).* Autocrine activation of an osteopontin-CD44–Rac pathway enhances invasion and transformation by H-RasV12. *Oncogene* **24**, 489–501.

[44] Sodek, Ganss, McKee (2000). Osteopontin. Crit Rev Oral Biol Med. **11**, 279–303.

[45] Prince, Dickie, Krumdieck (1991). Osteopontin, a substrate for transglutaminase and factor XIII activity. Biochem Biophys Res Commun. **177**, 1205–1210.

[46] Sørensen, Rasmussen, Møller, Jensen, Hørup, Petersen (1994). Localization of transglutaminase-reactive glutamine residues in bovine osteopontin. J Biochem. **304**, 13–16.

[47] Kaartinen, El-Maadawy, Räsänen, McKee (2002). Transglutaminase and its substrates in bone. J Bone Miner Res. **12**, 2161–2173.

[48] Uversky (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739–756

[49] Uversky *et al* (2000). Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427

[50] Williamson, M.P. (1994) The structure and function of proline-rich regions in proteins. *Biochem. J.* 297, 249–260

[51] Dyson, H. J. & Wright, P. E. (1998) Equilibrium NMR studies of unfolded and partially folded proteins. *Nat. Struct. Biol.* **5**, Suppl., 499–503.

[52] Lietzow, Jamin, Dyson, Wright (2002). Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **322,** 655–662.

[53] Klein-Seetharaman, Oikawa, Grimshaw, Wirmer, Duchardt, Ueda, Imoto, Smith, Dobson, Schwalbe (2002) Long-range interactions within a nonnative protein. *Science* **295,** 1719–1722.

[54] Shortle, Ackerman (2001). Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea. Science. **293,** 487–489.

[55] Mohana-Borges, Goto, Kroon, Dyson, Wright (2004). Structural Characterization of Unfolded States of Apomyoglobin using Residual Dipolar Couplings. J. Mol. Biol. **340**, 1131–1142.

[56] Robert Konrat (2009). The protein meta-structure: a novel concept for chemical and molecular biology. Cell. Mol. Life Sci.

[57] Jones, Wilkins, Smith, Dobson (1997). Characterisation of protein unfolding by NMR diffusion measurements. Journal of Biomolecular NMR. **10**, 199–203

[58] Christian, Kazanecki, Uzwiak, David, Denhardt (2007). Control of osteopontin signaling and function by post-translational phosphorylation and protein folding. Journal of Biochemistry. **102**, 912-924

[59] He, Mirza, Weber (2006). An osteopontin splice variant induces anchorage independence in human breast cancer cells. Oncogene. **25**, 2192-2202

9 Supplementary Information

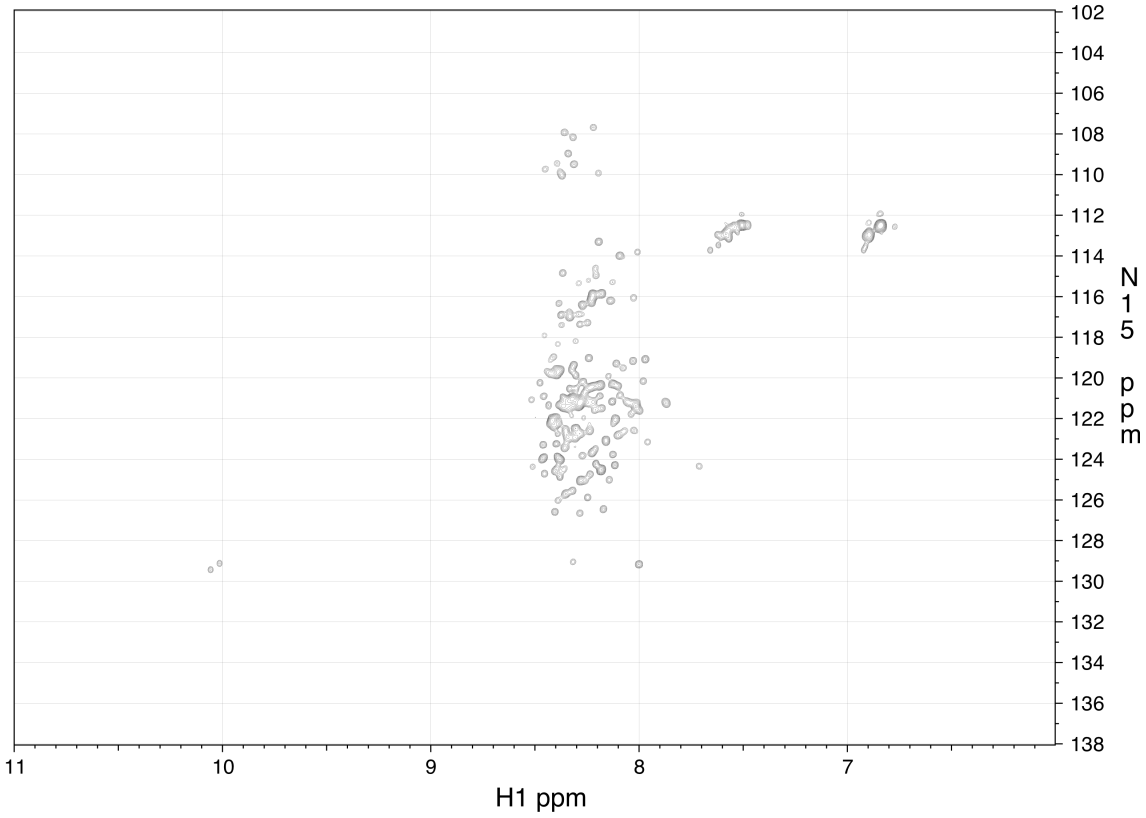


Figure 26: OPN220 ^{15}N-$^{1}H^{N}$ HSQC

10 Curriculum Vitae

Personal Information	Name: Gerald Platzer Nationality: Austria Date of Birth: 27.09.1982 Place of Birth: Graz Maritial status: single		
Education	1989 - 1993	Primary School	Weiz
	1993 - 1997	Highschool BG/BRG	Weiz
	1997 – 2002	Higher Technical School for Engineering Department for Electrical Engineering	Weiz
	2003 – 2009	University of Vienna Studies of Molecular Biology	Vienna
	2008-2009	MFPL – Diploma Thesis on the intrinsically unfolded protein Osteopontin.	Vienna
Internships	Feb 2007	MFPL – Group Hirt Cell Signalling	Vienna
	Aug,Sep 2007	University of Vienna – Group Stuessy Genetic Mechanisms of Evolution	Vienna
	Aug,Sep 2008	MFPL – Group Schlöglhofer Chromosome Biology	Vienna
Experience	Summer 2000-2007	Siemens 3D-Engineering and Design	Weiz
Skills	▪ Languages: English fluent, oral and written Spanish: basic knowledge Other Skills: ▪ PADI Open Water Diver ▪ Trained Paramedic (Red Cross & Austrian Army) ▪ Driving License		
Special Qualifications	2000 ▪ Unigraphics Engineer	VATech Hydro	Weiz

Vienna, 30.November 2009