BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

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

Gerald Platzer,[†] Andreas Schedlbauer,[†] Angela Chemelli,[‡] Przemyslaw Ozdowy,[§] Nicolas Coudevylle,[†] Renate Auer,[†] Georg Kontaxis,[†] Markus Hartl,[§] Andrew J. Miles,[∥] B. A. Wallace,[∥] Otto Glatter,[‡] Klaus Bister,[§] and Robert Konrat*[,†]

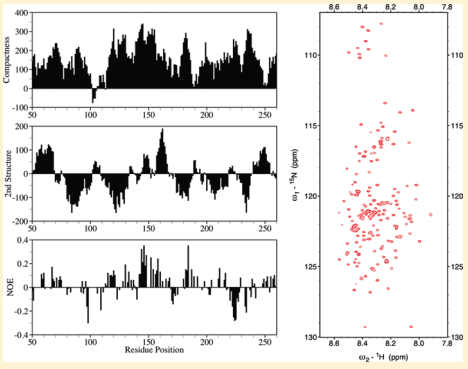
[†]Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Vienna Biocenter Campus 5, A-1030 Vienna, Austria

[‡]Institute of Chemistry, Physical Chemistry, University of Graz, Heinrichstrasse 28, A-8010 Graz, Austria

[§]Institute of Biochemistry and Center for Molecular Biosciences (CMBI), University of Innsbruck, Peter-Mayr-Strasse 1a, A-6020 Innsbruck, Austria

[∥]Department of Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, U.K.

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



Osteopontin (OPN), a secreted, sialic acid-rich, chemokine-like protein, is a member of the SIBLING family (small integrin-binding ligand N-linked glycoproteins).[1] OPN and other members of this protein family like bone sialoprotein, dentin matrix protein 1, dentin sialophosphoprotein, and matrix extracellular phosphoglycoprotein play important roles in many stages of cancer progression, as identified by gene transfer experiments.[2,3] OPN and most of the other family members are strongly overexpressed in several cancer cell lines, and their abundance directly correlates with the tumor grade.[4] For example, elevated expression levels of OPN were found in avian fibroblasts cooperatively transformed by the *v-myc* and *v-mil(raf)* oncogenes encoded by the acute leukemia and carcinoma virus MH2.[5] Promoter analysis revealed that the *OPN* gene is a direct transcriptional target gene of the activator protein-1 (AP-1) complex, mediated by a nonconsensus AP-1 binding site.[5] Overexpression of *OPN* results in an increased level of malignancy, whereas RNA silencing using appropriate antisense oligonucleotides leads to less malignant phenotypes.[6] A hallmark of OPN's biological functionality is the controlled interaction with a diverse set of receptor proteins.[7] OPN exerts its regulatory effects predominantly by interaction with integrin and CD44 receptors.[7] Additionally, OPN and other members of the SIBLING family have the capacity to bind and to activate matrix metalloproteases (MMPs).[1] The roles of OPN in diverse biological processes like atherosclerosis, bone remodeling, angiogenesis, wound healing, and tissue injuries have been reported.[7−9] The involvement of OPN in certain diseases such as arterial hyperplasia, myocardial necrosis, renal fibrosis, and autoimmune diseases is well-documented.[10] Of utmost medical importance is the fact that OPN is implicated in the pathology of diverse physiological processes as well as in tumor progression and metastasis in various types of cancer.[1,7]

Received: March 2, 2011
Revised: May 24, 2011
Published: May 24, 2011

OPN is classified as an arginine-glycine-aspartate (RGD)-containing glycoprotein predominantly found in kidney, brain, macrophages, vascular smooth muscle, and epithelial cells. It is an acidic protein, rich in aspartic and glutamic acids, and highly glycosylated, having an average molecular mass of ~44 kDa. The biological functionality of OPN is regulated by extensive phosphorylation.[11] It has a protease cleavage site separating two protein−protein interaction subdomains, an integrin-binding motif comprising the RGD sequence, and a C-terminal CD44 receptor binding site. However, the in vivo relevance of OPN activation by thrombin cleavage is still unclear. Finally, OPN also serves as a substrate for both liver transglutaminase and plasma transglutaminase factor XIIIa. Comprehensive overviews of OPN's biological functions of OPN have been published in recent years.[2,3,7]

Despite the significant biological and biomedical relevance of OPN, detailed information about its solution structural properties is still largely missing. Some preliminary insight was provided by recent NMR studies, where local structuring of OPN was demonstrated.[12,13] Here we have investigated the solution structure of a recombinant OPN protein by an integrated structural biology approach employing multidimensional nuclear magnetic resonance (NMR) spectroscopy, synchrotron radiation circular dichroism (SRCD) spectroscopy, and small-angle X-ray scattering (SAXS). We show that OPN is devoid of any significant tertiary structure, largely lacking secondary structures, and thus represents another member of the growing family of intrinsically disordered and conformationally flexible proteins. The conformational ensemble of quail OPN was further investigated using residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs). For the identification of suitable sites of spin-label attachment, we employed our recently developed meta-structure approach.[14] NMR spectroscopy was also used to investigate putative calcium ion and heparin binding sites in the OPN protein. No evidence of calcium binding was obtained, but the binding site of heparin was mapped to the central integrin-binding domain of OPN comprising the RGD motif. On the basis of multidimensional NMR spectroscopy, we provide evidence that OPN exists as a conformationally flexible polypeptide chain in solution devoid of stabilizing long-range interactions. However, despite the fact that OPN does not form an energetically well-defined three-dimensional (3D) solution structure, it displays several local secondary structure elements with significantly reduced conformational flexibilities. Additionally, SAXS and PRE data indicate that although OPN exists in a rather extended form, there is evidence of a transient compaction of the polypeptide chain. This motional heterogeneity along the protein backbone may be of relevance for the biological function of OPN. It could allow adaptive processes for optimization and tuning of intermolecular interactions with diverse protein binding partners across the relevant interaction interfaces of OPN.

MATERIALS AND METHODS

Protein Expression. The recombinant quail OPN protein (OPN220) contains the 219 C-terminal amino acids of quail OPN and a vector-encoded start methionine.[5,12] *Escherichia coli* strain BL21(DE3) transformed with the pET11d-OPN220 expression vector described previously[12] was grown in LB medium overnight. The culture was diluted 1:100 in 50 mL of minimal medium, supplemented with 0.5 g of NaCl, 1 g of $^{15}NH_4Cl$, and 2 g of D-[^{13}C]glucose per 1000 mL, and the bacteria were incubated at 37 °C for 12−16 h. The 50 mL preculture was added to 1000 mL of minimal medium, and the bacteria were grown at 37 °C until the culture reached an absorbance at 600 nm of 0.7. Recombinant protein synthesis was induced by the addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and incubation was continued for 12 h at 28 °C. The cells were collected by centrifugation at 5000 rpm for 20 min and resuspended in 20 mL of ice-cold PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4) per liter of the original bacterial culture. Bacteria were lysed by sonication, and the cell lysate was cleared by centrifugation at 18000 rpm for 30 min; 20 mL portions of the supernatant containing the soluble protein fraction were adjusted to a $(NH_4)_2SO_4$ concentration of 50% and then clarified by centrifugation at 18000 rpm for 30 min. The precipitate containing most of the OPN recombinant protein was dialyzed against PBS buffer at 4 °C and loaded onto a Resource Q, 6 mL ion exchange column (GE Healthcare). Protein was eluted with a sodium chloride gradient generated by mixing PBS and high-salt PBS buffer (1 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4). The fractions containing most of the OPN protein were concentrated to 1.5 mL by centrifugation through an Amicon Ultra-15 centrifugal filter device 10K NMWL (Amicon) and loaded onto a Superdex 200 HiLoad 16/60 prep grade (GE Healthcare) gel filtration column equilibrated in phosphate buffer [50 mM NaCl and 50 mM sodium phosphate (pH 6.5)]. The final yield of homogeneous OPN was approximately 4 mg/L of bacterial culture. For NMR analysis, protein samples were concentrated to 0.5 mM.

Synchrotron Radiation Circular Dichroism (SRCD) Spectroscopy. Protein concentrations were determined by absorbance measurements (NanoDrop UV/vis spectrophotometer using the theoretical extinction coefficient calculated using the ProtParam tool at http://www.expasy.org). Samples containing 0.5 mM of the human form of the protein dissolved in phosphate buffer [50 mM NaCl and 50 mM sodium phosphate (pH 6.5)] were loaded into a demountable Suprasil quartz cell (Hellma UK Ltd.) with a path length of 0.0015 cm. The spectrum was measured at 25 °C over the wavelength range from 170 to 280 nm with a step size of 1 nm and a dwell time of 2 s. Data were processed using CDTool.[15] Three sample scans were averaged, and three baseline scans (consisting of buffer) were averaged and subtracted from the sample scans. The resulting spectrum was smoothed using a Savitsky-Golay filter, calibrated against a spectrum of camphorsulfonic acid,[16] and then converted to $\Delta\varepsilon$ units using a mean residue weight of 113.7. Secondary structure content was analyzed using the DichroWeb analysis server,[17] using reference data set 6,[18] which contains the spectra of five denatured proteins and hence produces a more accurate prediction when the protein contains disordered structure. The results from three different algorithms, CONTINLL,[19,20] SELCON3,[21] and CDSSTR,[21] were averaged.

NMR Spectroscopy. NMR spectra were recorded at 25 °C using Varian Inova 500 and 800 MHz and Varian Direct Drive 600 MHz spectrometers. Spectra were recorded in the States-TPPI/PFG sensitivity-enhanced mode[22] for quadrature detection with carrier frequencies for $^1H^N$ and ^{15}N of 4.73 and 120.0 ppm, respectively. The sample contained 0.5 mM recombinant quail OPN, 50 mM NaCl, and 50 mM sodium phosphate (pH 6.5) in a 90% H_2O/10% D_2O mixture. The suite of standard triple-resonance experiments performed for extraction of $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts included 3D HNCA, 3D HN(CO)CA, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO, and 3D HN(CA)CO.[23] In addition, 3D HNN and 3D HN(C)N experiments[24] were conducted to establish sequential connectivities

a)
```
  1 MKLTLLCLCF ISITAAWPVS KSKQHAISAS
 31 SEEKYDPRSH HTHRYHQDHV DSQSQEHLQQ
 61 TQNDLASLQQ THYSSEENAD VPEQPDFPDV
 91 PSKSQETVDD DDDDDNDSND TDESDEVFTD
121 FPTEAPVAPF NRGDNAGRGD SVAYGFRAKA
151 HVVKASKIRK AARKLIEDDA TTEDGDSQPA
181 GLWWPKESRE QNSRELPQHQ SVENDSRPKF
211 DSREVDGGDS KASAGVDSRE SQGSVPAVDA
241 SNQTLESAED AEDRHSIENN EVTR
```

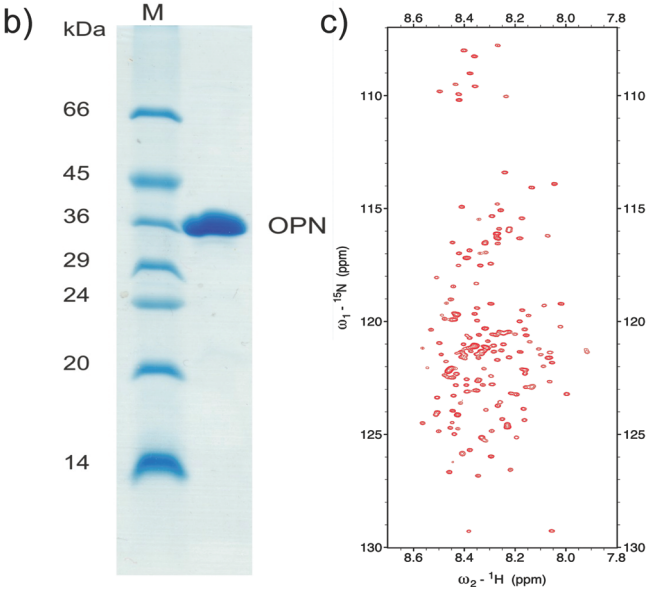


Figure 1. Purity of the 220-amino acid recombinant OPN (OPN220) monitored by sodium dodecyl sulfate−polyacrylamide (17%, w/v) gel electrophoresis and 2D $^{1}H-^{15}N$ heteronuclear single-quantum coherence (HSQC) spectroscopy. (a) Amino acid sequence of wild-type quail OPN. OPN220 contains the 219 carboxyl-terminal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

and to obtain a large extent of unambiguous backbone assignments. NMR spectra were processed and analyzed with NMRPipe[25] and SPARKY.[26] High-quality NMR data for OPN were obtained as shown by the PFG sensitivity-enhanced two-dimensional (2D) $^{1}H-^{15}N$ HSQC spectrum recorded at 800 MHz (cf. Figure 1c). A squared and 60° phase-shifted sine bell window function was applied in all dimensions for apodization. Time domain data were zero-filled to twice the data set size, prior to Fourier transformation. The applied pulse sequences are sensitivity-enhanced and use gradients for coherence selection and water suppression. ^{1}H chemical shifts were referenced as reported previously.[12] Heparin was purchased from Sigma Aldrich and used without further treatment.

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

^{15}N Relaxation. Dynamics information was obtained by measurement of backbone ^{15}N-T_1 and -T_2 relaxation times as described previously.[29] Relaxation parameters were measured at 25 °C employing a magnetic field strength that corresponds to a ^{1}H Larmor frequency of 599.89 MHz. The ^{15}N transverse relaxations experiments for evaluating T_2 were performed with Carr−Purcell−Meiboom−Gill (CPMG) delays of 0, 16, 32, 64, 128, 192, and 256 ms using a CPMG duty cycle delay of 0.5 ms. ^{15}N longitudinal relaxation T_1 measurements were obtained by employing delays of 0, 20.8, 41.6, 83.2, 208, 416, 832, 1248, and 1664 ms. The sequence of cross-peak intensities in this series of spectra was fit (using an extension script of SPARKY[26]) to a two-parameter exponential decay of the form $I(t) = A \exp[-(t/T)]$, where $I(t)$ represents the peak intensity and t the delay time. Random errors for T_1 and T_2 values were estimated by calculating a best fit for a set of perturbed heights 5000 times. The heights were perturbed by computer-simulated noise with a Gaussian distribution of zero mean and with variance equal to the root-mean-square deviation of the original heights from the original best fit. The spread in the best fit values was used as a measure of the experimental random error.

Heteronuclear steady-state NOE $^{15}N-\{^{1}H^{N}\}$ attenuation factors were derived from the I_{NOE}/I_{nonNOE} ratio, where I_{NOE} and I_{nonNOE} denote the peak intensities in the experiments with and without proton saturation, respectively.[29] In the case of spectra with nonsaturation, a net relaxation delay of 5 s was employed whereas a relaxation delay of 2 s prior to a 3 s proton presaturation period was applied for the NOE spectra. Acquisition parameters were identical to those of the PFG sensitivity-enhanced 2D $^{1}H-^{15}N$ HSQC experiment except that the number of transients per t_1 value was set to 32. ^{15}N CPMG dispersion experiments were performed as described by Kay and co-workers.[30]

Measurement of Hydrodynamic Radii. Pulse field gradient (PFG) NMR diffusion measurements were performed on a Varian Direct Drive spectrometer equipped with a z-gradient probe operating at 600 MHz. The sample was supplemented with 0.2% dioxane as an internal reference. The PG-SLED (pulse gradient stimulated echo longitudinal encode-decode) sequence was used as described by Wilkins et al.,[31] to which a final WATERGATE module[32] was added. The applied diffusion gradient time (δ) was 4.5 ms, and the echo time (Δ) was 100 ms. Fifty experiments were conducted with increasing diffusion gradient strength. The data were processed using NMRPipe.[25] The decay of the protein aliphatic proton signals was fitted to a single Gaussian using the DOSY module of NMRPipe to determine the diffusion constant of the protein (d_{prot}). The same procedure was applied to the dioxane resonance to determine the diffusion constant of dioxane (d_{diox}). The hydrodynamic radius of the protein (R_h^{prot}) was deduced using the following equation:

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

assuming that R_h^{diox} (hydrodynamic radius of dioxane) is 2.12 Å.[31]

dx.doi.org/10.1021/bi200291e |*Biochemistry* 2011, 50, 6113–6124

Residual Dipolar Couplings (RDCs). To obtain $^{15}N-^{1}H^{N}$ couplings, a $^{1}H-^{15}N$ HSQC spectrum without proton decoupling during nitrogen evolution was recorded. The resulting signal splitting leads to a doubling of peaks, which might result in a crowded spectrum with signal overlap. To circumvent this problem, we recorded two separate spectra. One contains the inphase component of the coupled magnetization and the other the antiphase component (IPAP). Both spectra were added and subtracted from each other to separately yield the upfield and downfield components of the doublet. Splittings were measured under isotropic (J_{HN}) and aligned ($J_{HN} + D_{HN}$) conditions to isolate the RDC contribution.[33]

Paramagnetic Relaxation Enhancements (PREs). Paramagnetic relaxation enhancements by nitroxide spin-labels [4-(2-iodoacetamido)-TEMPO] in single-cysteine-mutant OPN variants were measured as intensity ratios of cross-peaks in 2D $^{1}H-^{15}N$ HSQC NMR spectra in the presence and absence of the nitroxide radical. The individual cross-peak intensities were normalized using the most intense peaks in the two data sets. The mutant OPN proteins were expressed and purified from plasmids that have been mutagenized using the original pET11d-OPN220 construct as a template and the QuickChange II site-directed mutagenesis kit (Stratagene). The following complementary (f, forward; r, reverse) primers were used: Opn220-c108-f, 5′-GATGATGACAATGATTGCAATGACACCGATGAG-3′; Opn220-c108-r, 5′-CTCATCGGTGTCATTGCAATCATTGTCATCATC-3′; Opn220-c188-f, 5′-TGGTGGCCCAAGGAGTGCCGCGAACAGAACAGC3′; Opn220-c188-r, 5′-GCTGTTCTGTTCGCGGCACTCCTTGGGCCACCA-3′.

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

Small-Angle X-ray Scattering. The SAXS equipment consisted of a slit-geometry camera with high flux and low 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 (Princeton Instruments/Roper Scientifics, Trenton, NJ) was used to record the 2D scattering pattern. The temperature of the capillary in the metallic sample holder was controlled by a Peltier element. Measurements were taken at 25 °C. The measuring times were 3 × 10 min for all samples. This allowed for the proper subtraction of cosmic rays that appear when using a CCD camera. The images were then integrated into the one-dimensional scattering function $I(q)$. This one-dimensional scattering curve is a function of the magnitude of the scattering vector $q = (4\pi/\lambda)\sin(\theta/2)$, where θ is the total scattering angle. All the intensities were transmission-calibrated by normalizing the attenuated primary intensity at q = 0 to unity and were corrected for the background scattering from the capillary and the buffer solution. The absolute scale calibration was achieved by using water as a secondary standard.[34]

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

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

where r is the distance between two scattering centers within the particle.

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

RESULTS

NMR Spectroscopic Analysis of OPN and Resonance Assignments. Because of inefficient bacterial expression of the full-length quail OPN protein, a 220-amino acid recombinant OPN derivative lacking amino acid residues 2−45 of the 264-amino acid quail OPN was expressed and purified as described previously.[12] It should be noted that all amino acid positions addressed in this report refer to the numbering of the full-length 264-amino acid protein. The purity of the protein sample was verified by SDS−PAGE and NMR spectroscopy (Figure 1). The acidic OPN220 (M_r = 24311; pI = 3.94) migrates with an apparent molecular weight of 35000 (Figure 1b). This anomalous electrophoretic behavior in SDS−PAGE is typical for highly acidic proteins.[5] Electrospray ionization (ESI) mass spectrometry was performed to confirm the identity of the purified recombinant protein (data not shown). The $^{1}H-^{15}N$ HSQC spectrum shows a distinct separation of the cross-peaks into regions typical of residue random coil positions (Figure 1c). There are resonance clusters for glycines, serines/threonines, and all other backbone amides. One underlying problem in the assignment of an intrinsically unfolded protein like OPN is the rather poor extent of backbone amide $^{1}H^{N}$, $^{13}C\alpha$, and $^{13}C\beta$ chemical shift dispersion compared to that of a protein molecule in the folded state. Ring current and other dipolar interactions in unfolded states are averaged out by conformational fluctuations leading to chemical shifts close to random coil values. Consequently, the $^{1}H^{N}$, $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts for the same residue type tend to degenerate. Furthermore, considering the relatively large molecular size of the 220-amino acid OPN protein, the difficulty of assignment becomes unfavorable when employing only a set of standard experiments such as 3D HNCA, 3D HN(CO)CA, 3D HNCO, 3D HN(CA)CO, 3D HNCACB, and 3D CBCA(CO)NH, which are routinely used for backbone assignment of folded biological molecules.[23] Problems of spin system overlap and degeneracy can be partly overcome by the incorporation of sequence-dependent amide nitrogen ^{15}N chemical shifts. This was exploited by applying HNN and HN(C)N triple-resonance experiments that sequentially correlate $^{1}H^{N}$ and ^{15}N resonance chemical shifts.[24] In total, 181 of 207 non-proline $^{1}H^{N}$ and ^{15}N backbone chemical shifts were assigned (87.4%). $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts were allocated for 195 residues (89.0%). For the N-terminal stretch containing residues His46−His49, no signal could be found presumably because of conformational dynamics as indicated by the very low signal intensity of residue Val50 in the HSQC spectra. Only the two initial residues of a polyaspartic stretch from Asp99 to Asp105 could be unambiguously assigned. The 17 remaining unassigned residues included Gln53−His57, Asn63, Asp64, Glu77, Asp89,

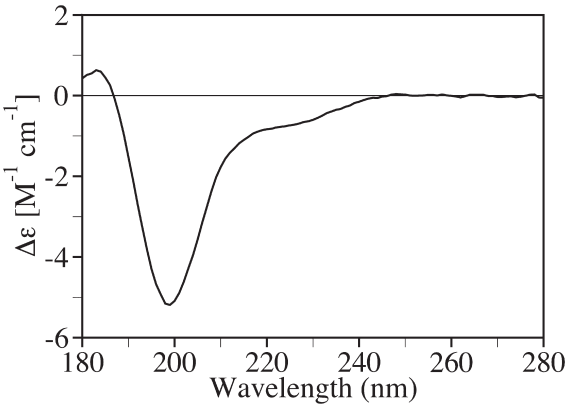


Figure 2. Synchrotron radiation circular dichroism (SRCD) spectra of OPN in 50 mM NaCl and 50 mM sodium phosphate (pH 6.5). Deconvolution analyses indicate the following secondary structure components are present: 5% α-helix, 13% β-sheet, and 82% other or unordered.

Val90, Glu116 Gln178, His199, Arg207, Arg213, Asn260, and Glu261, which could not be allocated unambiguously because of intense spin system overlap in both $^{1}H^{N}$ and ^{15}N dimensions. The assignments have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) as BMRB entry 15519.[12]

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

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

Secondary Chemical Shifts. The differences between the observed chemical shifts of nucleus spin types and their random coil values assigned in unfolded conformations serve as a valuable tool for evaluating the residual preferences for protein secondary structure. Figure 3 shows the sequence-corrected secondary $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts for OPN. In α-helices, the $^{13}C\alpha$ values appear to be shifted downward relative to their random coil values, whereas the $^{13}C\beta$ values are shifted upward, leading to opposite signed effects of both nuclear spins. In β-strands, the

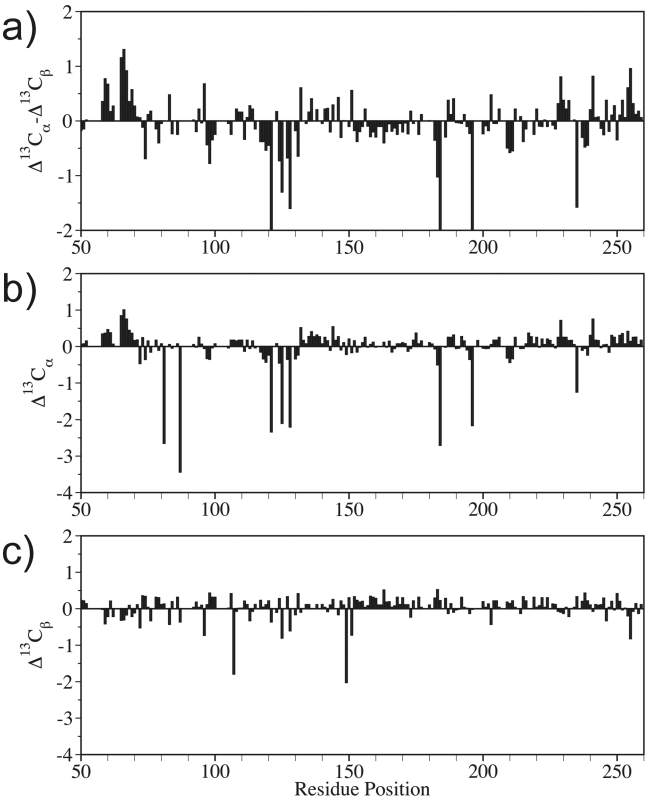


Figure 3. Residue plots of secondary ($^{13}C\alpha-^{13}C\beta$), $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts. Shown are $\Delta^{13}C\alpha - \Delta^{13}C\beta$ (a), $\Delta^{13}C\alpha$ (b), and $\Delta^{13}C\beta$ (c) chemical shift deviations from random coil values in OPN. Random coil values were derived from ref 63. Significant α-helical propensities were found for residues 60−70 and 130−150, whereas residues 120−130 display extended (β-strand-like) conformations. Additionally, small residue stretches in the C-terminus of OPN also display non-zero α-helical probabilities.

reverse applies. Overall, the majority of the residues display chemical shift values that are typical for random coils. However, for some residues, small secondary shifts indicating a slight tendency toward a α-helical conformation are apparent, like Leu58 to Leu68, or the regions around Ala150 and His255. Slight residual preferences for a β-strand could be inferred from the chemical shifts for regions around residues Asp80 and Ala125. Additionally, we also used the TALOS database system to probe the existence of local secondary structure elements.[38] Overall, only a small number of residues converged to unique solutions in Ramachandran space, thus underscoring the finding that OPN displays significant conformational flexibility. Pronounced tendencies to adopt (non-random coil) β-strand conformations were found for only a small number of residues: Thr71, Thr97, Thr119, Ser141, the fragment of residues Val152−Lys154, Lys157, and Ile158.

NMR Relaxation Rates and Solution Motional Behavior of OPN. Measurement of NMR relaxation rates provides a window on protein dynamics over a broad range of time scales.[23] The longitudinal ^{15}N-T_1 values, transverse ^{15}N-T_2 values, and heteronuclear steady-state NOE $^{15}N-\{^{1}H\}$ attenuation factors are sensitive to dynamics on the picosecond to nanosecond time scale, whereas ^{15}N-T_2 can also be affected by conformational or chemical exchange processes on the millisecond to microsecond time scale. Despite the rather small chemical shift dispersion of OPN in the $^{1}H-^{15}N$ HQSC spectrum, ^{15}N relaxation data could

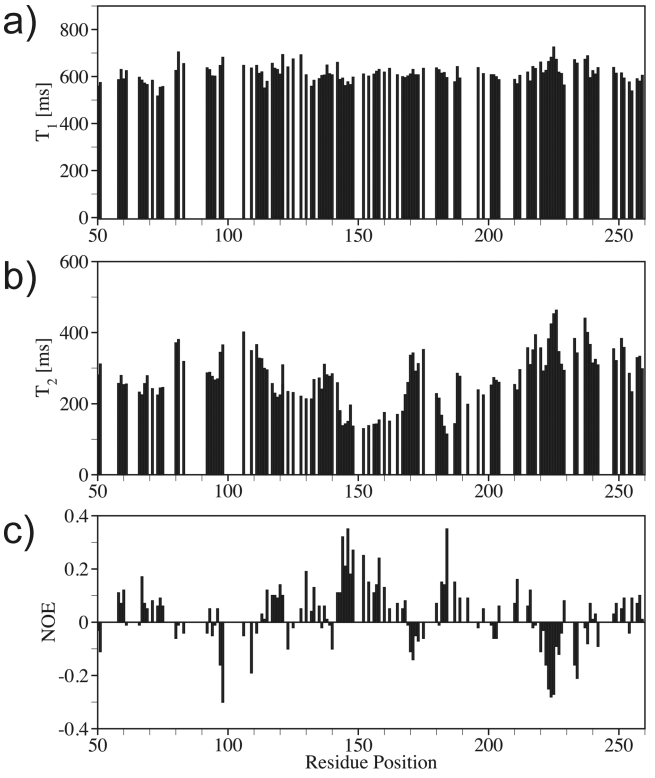


Figure 4. ^{15}N relaxation data of OPN measured at 600 MHz. ^{15}N-T_1 (a), ^{15}N-T_2 (b), and steady-state ^{15}N−{^{1}H} heteronuclear NOE (c) plotted as a function of residue position.

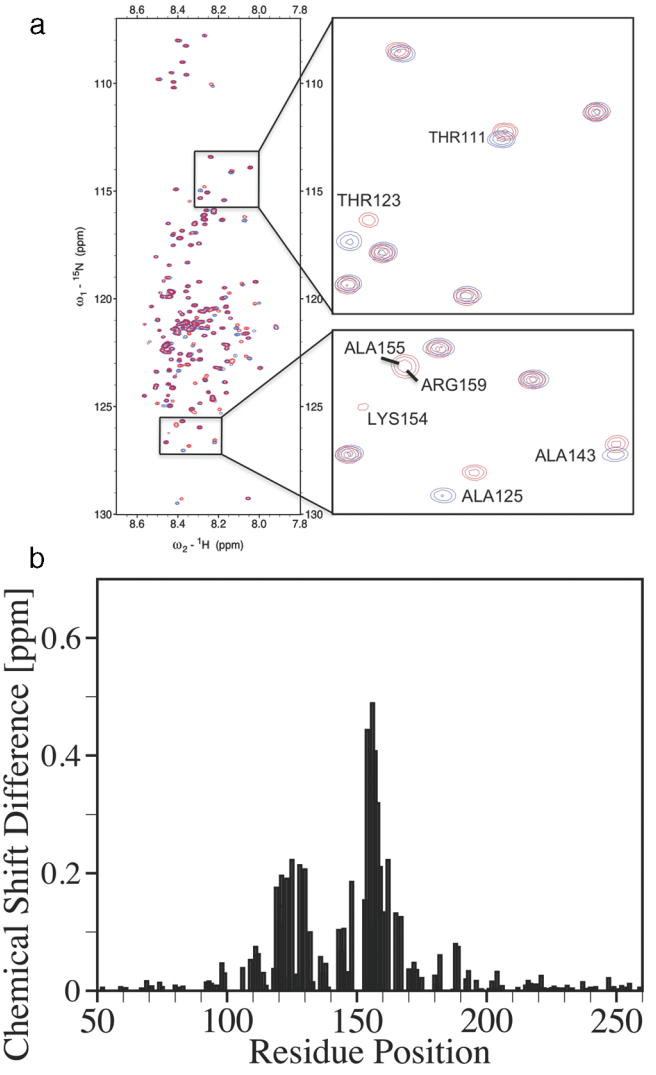


Figure 5. Heparin binding site of OPN. (a) Mapping of the heparin binding site was carried out by monitoring ^{15}N−^{1}H chemical shift changes of OPN via 2D ^{1}H−^{15}N HSQC spectroscopy. (b) Chemical shift differences between the heparin-free and -bound form of OPN220 plotted as a function of residue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

be obtained for 122 of 183 assigned non-proline residues at 298 K and 600 MHz. As indicated in Figure 4a, ^{15}N-T_1 values exhibit a strictly monotonous and featureless distribution along the amino acid sequence except at the C-terminal end, with an average value of 619 ms. The residue plot of the experimental ^{15}N-T_2 values (Figure 4b) shows several sequential variations ranging from 114 to 747 ms. This is in strong contrast to an expected parabolic form with a ^{15}N-T_2 minimum in the middle of the sequence arising from a complete random structure without any structural preferences. Overall, there is good agreement between secondary chemical shifts indicating secondary structure formation and motional restrictions probed by ^{15}N-T_2 relaxation. Interestingly, the Leu182-Trp183-Trp184 stretch also appears to be constrained presumably because of the bulky and sterically demanding indole side chains. These findings are in good agreement with the obtained steady-state NOE ^{15}N−{^{1}H} attenuation factors (Figure 4c) that are found to be generally higher and strictly positive in these ranges. The highest values (0.2−0.37) were observed in the middle of the sequence and for residue Trp184 (0.35). Extensive patches of negative attenuation factors were apparent from residue Ala170 to Gly175 and more pronounced along a nine-residue stretch comprising residues Gly217−Gly225, a motif consisting of residues of small side chains (three glycines, two alanines, and two serines). In addition, these parts display higher ^{15}N-T_2 values of 350−450 ms. The total average of ^{15}N-T_2 is 280 ms. The intramolecular dynamics of OPN on a microsecond to millisecond time scale was investigated by state-of-the-art ^{15}N CPMG dispersion experiments. However, analysis of the effective relaxation rate R_2 versus the CPMG frequency (ν_{CPMG}) profile revealed flat dispersion profiles for all residues resolved in the 2D ^{1}H−^{15}N HSQC spectra indicating a lack of microsecond to millisecond time scale motions in OPN.

Mapping of the Heparin Interaction Site. It has been reported that OPN binds to heparin, although no clear experimental evidence was given.[39,40] To provide more specific information about the binding site and the mode of binding, we applied ^{1}H−^{15}N HSQC spectroscopy. Figure 5a shows an overlay of HSQC spectra of unligated OPN and after binding to heparin. The chemical shift differences between the heparin-free and -bound forms of OPN220 plotted as a function of residue position are shown in Figure 5b. There is evidence of a specific interaction between OPN and heparin. Although the majority of peaks were unchanged, several residues experienced significant chemical shift changes upon heparin binding. Overall, the magnitudes of chemical shift changes are small, presumably because heparin lacks aromatic ring systems and because ^{1}H−^{15}N HSQC spectra report on the protein backbone and not on side chains, which are in direct contact with the ligand.

Mapping of the residues displaying frequency changes upon binding of heparin on the protein backbone clearly reveals a distinct binding side near the central integrin-binding domain (residues 117−130 and 142−167). Furthermore, OPN is an inhibitor of hydroxyapatite (HA)[41−43] and calcium oxalate (CaOx) crystallization.[44] To elucidate the effects of Ca^{2+} on OPN, we performed $^{1}H-^{15}N$ HSQC spectroscopy. The $^{1}H-^{15}N$ HSQC spectra remained nearly unchanged even at a 40:1 molar excess of calcium ions over OPN (data not shown), thus further corroborating the previous results showing that post-translational modification (phosphorylation) is an important prerequisite for effective Ca^{2+} binding of OPN.[40,45]

Meta-Structure Analysis. The meta-structure concept was recently introduced as a novel theoretical framework for protein sequence analysis.[14] The protein meta-structure information provides two quantitative parameters, compactness and local secondary structure, describing the topology of a protein. The local residue-specific compactness value (C_i) is a quantitative parameter describing the structural complexity of an individual residue in the context of the 3D protein fold.[14,46] Residues that are on average deeply buried in the interior of a structure exhibit large C_i values, whereas small C_i values are characteristic of flexible loop regions and/or intrinsically unfolded segments of the polypeptide chain. The predicted local secondary structure parameter S_i referring to the local secondary structure is defined in analogy to the well-established $^{13}C\alpha$ secondary shifts. Positive S_i values are indicative of α-helix and negative S_i values the presence of an extended conformation. For the sequence-based prediction of protein disorder, several bioinformatic approaches that are primarily based on neural networks have been developed.[47−53] Meta-structure analysis differs from these bioinformatics approaches by providing quantitative information defined on a per residue basis, which can be used to analyze details of protein folds on a residue resolution level. To evaluate the performance and usefulness of the meta-structure approach for the characterization of IDPs, we compared our data to disorder prediction results obtained with DISOPRED2.[52] The sequence-derived compactness data for OPN and the comparison with DISOPRED2 results as well as steady-state $^{15}N-\{^{1}H\}$ heteronuclear NOE are shown in Figure 6. For the calculation of compactness values, the 264-amino acid full-length quail OPN was used. For OPN, the average residue compactness (ARC) value of ~182 was observed, which is significantly lower than the average protein database (PDB) value of 300, suggesting that OPN does not form a stable tertiary fold in solution. Figure 6 shows a comparison of DISOPRED2 results (a) with predicted local compactness values (Figure 6b) and steady-state $^{15}N-\{^{1}H\}$ heteronuclear NOEs (Figure 6c). The predicted local secondary structural features determined by meta-structure analysis (Figure 7) correlate reasonably well with the NMR-based identification of secondary structures. The locations of the N-terminal α-helix (Gln59−Ser74) and of the two extended regions (Glu77−Phe87 and Val117−Arg132) were correctly reproduced by the meta-structure parameters. For comparison, the primary sequence of OPN was also subjected to the well-established disorder prediction server DISOPRED2.[52] The obtained prediction results indicate that OPN exists as an intrinsically unstructured protein in solution. Specifically, in the C-terminal region of OPN beyond residue position 190, disorder probability values larger than 0.6 were obtained, whereas in the N-terminus, disorder probabilities were consistently smaller, ranging from 0.005 at residue 125 to 0.42 at residue 28. For the most compact segment of residues 115−190, the program consistently provides minimum

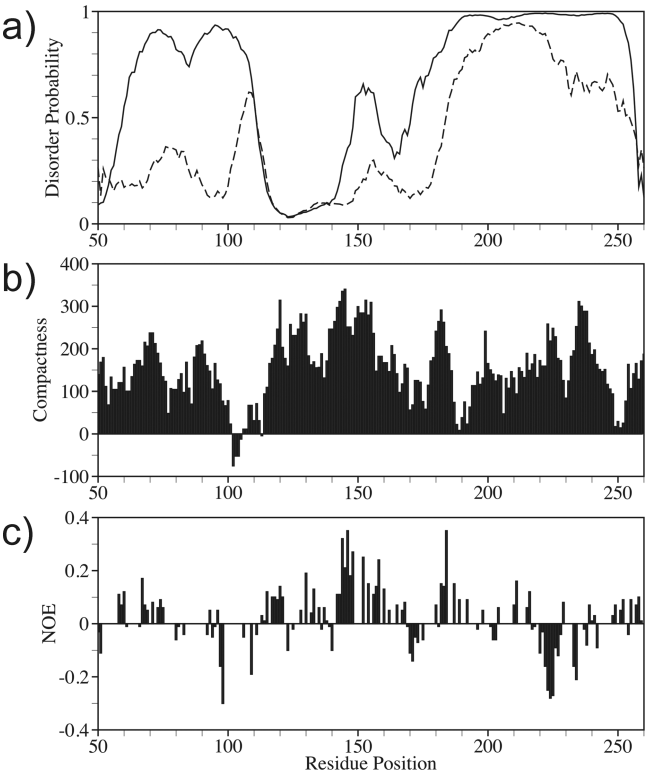


Figure 6. Meta-structure analysis and comparison with DISOPRED2 and NMR relaxation data. DISOPRED2-derived disorder propensities (a) and meta-structural parameter compactness (b) and steady-state $^{15}N-\{^{1}H\}$ heteronuclear NOE (c) plotted as a function of residue position. Large compactness values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

disorder probabilities of <0.1. The DISOPRED2 secondary structure prediction expressed in *psipred* values also provides evidence of the presence of small local secondary structure elements in OPN. While the α-helical *psipred* values identified for residues 165−170 are in good agreement with both our meta-structure prediction and NMR secondary shift data, the helical segment of residues 55−70 both predicted by meta-structure analysis and experimentally observed by NMR was not identified by DISOPRED2. β-Sheets were predicted by DISOPRED2 for residues 119−121 and 143−148. For residues 119−127, the existence of a β-sheet was also indicated by meta-structure analysis and NMR spectroscopy. In contrast, the stretch of residues 143−148 did not display significantly extended conformations. Overall, meta-structure analysis provided reliable information about the structural preferences of OPN in solution and about the presence and location of secondary structure elements that nicely correlate with experimental NMR data. Interestingly, residues displaying larger compactness values (centered around residues 125, 150, and 180) consistently exhibited reduced local backbone mobility as evidenced by ^{15}N relaxation data. This observation underscores the reliability of the predicted

compactness values and suggests fruitful applications to experimental NMR spectroscopy of IUPs (identification of spin-label attachment sites) (see below).

Residual Dipolar Couplings and Paramagnetic Relaxation Enhancements. Recently, it has been shown that residual dipolar couplings (RDCs) and paramagnetic relaxation enhancement data (PRE) are versatile spectral probes for the characterization of structural dynamics of intrinsically or natively unfolded proteins.[54−56] RDCs report on time- and ensemble-averaged conformations of polypeptide chains and have been successfully used to study unfolded proteins in solution.[55,57,58] The partial alignment of molecules allows the detection of small degrees of backbone compaction. Measurements of $^{1}H-^{15}N$ backbone RDCs in aligned *pf*-1 phage media (10 mg/mL) probe all accessible vector bond orientations up to tens of microseconds.

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

PREs report on long-range interactions and thus provide valuable information about transient contact formation in unfolded proteins.[56,58,59] Despite the ease of experimental implementation, one of the inherent problems is the choice of suitable spin-label sites, e.g., sites of Cys mutations. Here we have used meta-structure-derived compactness data to identify suitable sites of spin-label attachment. Figure 8a shows the compactness as a function of residue position. The sites of spin-label attachment were selected on the basis of minimal compactness value, as for these regions tight side chain interactions or packing can presumably be neglected. Panels b and c of Figure 8 show PRE data obtained for OPN mutants S108C and S188C, respectively. It can be seen that ^{15}N paramagnetic relaxation enhancements are not limited to only neighboring residues but also appear in regions distant from the affected mutation sites. This indicates some degree of compaction and transient contact formation. Most importantly, residues with significant PREs are observed in regions for which both positive heteronuclear NOEs were observed and large compactness values were predicted like residues 115−160. The pulse field gradient (PFG) NMR diffusion data and the large hydrodynamic radius (see below) indicate that OPN exists in solution as a largely extended polypeptide chain with some local compaction in the central domain that encompasses the integrin-binding domain. Independent NMR evidence corroborating this collapsed state was recently provided using paramagnetic relaxation enhancement data where transient interactions between the N- and C-terminal domains were observed.[13]

Measurement of Hydrodynamic Radii. We determined by PFG NMR diffusion that the hydrodynamic radius R_h of OPN is

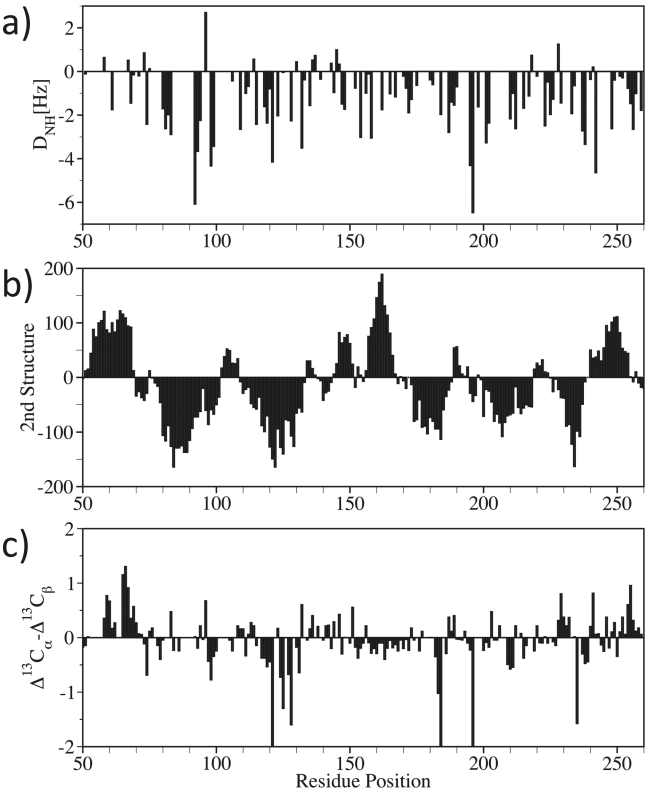


Figure 7. Secondary structure of OPN. Comparison between $^{1}H^{N}-^{15}N$ residual dipolar couplings (RDCs) (a), meta-structure-derived secondary structure (b), and $^{13}C\alpha-^{13}C\beta$ NMR secondary shifts (c). Meta-structure-derived local secondary structure values are defined in a similar manner compared to NMR $^{13}C\alpha$ secondary shifts. Positive values are indicative of α-helical regions, whereas continuously negative values are typical for extended or β-strand regions. OPN was aligned in a buffer containing 50 mM NaCl and 50 mM sodium phosphate (pH 6.5) supplemented with a colloidal *pf*-1 phage suspension (10 mg/mL). Note that the RDCs are plotted according to convention on an inverted axis where negative values are shown as positive ones.

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

Independent experimental data for the solution behavior of OPN were provided by a small-angle X-ray scattering analysis. The one-dimensional scattering curves were further treated using GIFT.[34−36] Because of the low protein concentration, the contribution of protein interactions is negligible. Furthermore, intermolecular electrostatic interactions caused by ions in the solution buffer also do not contribute to the scattering curve. The scattering was therefore only related to the overall shape, the electron density, and its distribution within the protein. The pair distance distribution function (PDDF) could be obtained by IFT. This function can be visualized as a histogram of distances present in the protein. The shape of the PDDF can be evaluated

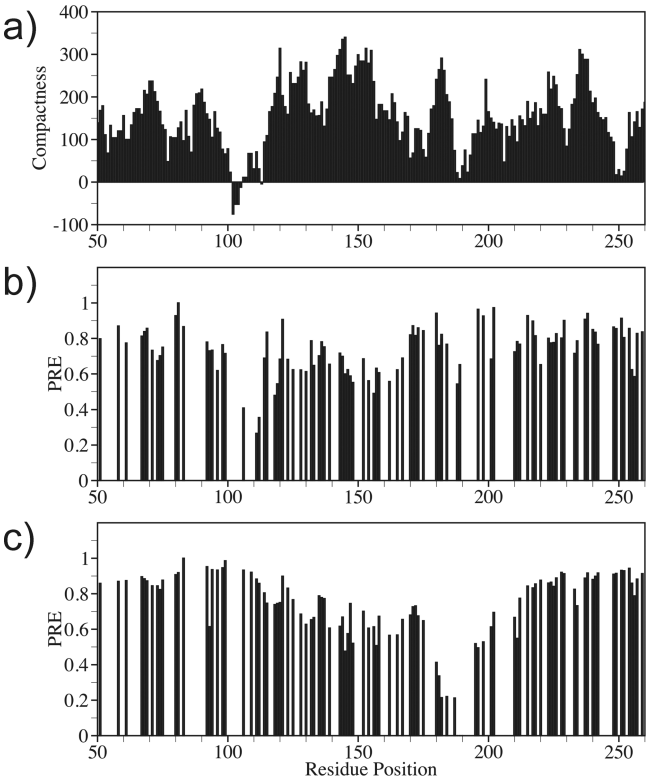


Figure 8. Residue plot of protein compactness and paramagnetic relaxation enhancement (PRE) for OPN. (a) The protein compactness data were used to identify possible sites for spin-label attachment. Residues displaying small compactness values are found in exposed and/or conformationally flexible parts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

to gain information about the structure of the protein. For a prototypical Gaussian coil polymer, the PDDF shows a linear increase at short distances and a slow decay at longer distances. Figure 9 shows the PDDF extracted from the experimental scattering curve of OPN. While the linear increase could be clearly observed in the PDDF of the protein, the decrease in the PDDF function is flatter than that one would expect for a Gaussian coil. The maximum dimension of the protein could be estimated to ∼210 Å. Furthermore, a shoulder in the function at ∼15 Å was observed. It is important to mention that the scattering curve represents a sum of the scattering, which is caused by all structures appearing in the measured volume during the measurement time. The extracted PDDF clearly indicates that OPN comprises a heterogeneous conformational space containing both extended coil and compact conformation with distinct side chain interactions. For comparison, Figure 9 also shows the PDDF curve of the stably folded lipocalin protein Q83[60] that strikingly differs from that of OPN. The average gyration radius of OPN was determined to be 51.3 Å. For comparison, the average gyration radius of Q83 was determined to be 20.7 Å, which is in very good agreement with the 3D structural data obtained by NMR spectroscopy.[60] The data obtained for OPN suggest that the analysis of PDDFs provides valuable information about the conformational ensembles of intrinsically disordered proteins in

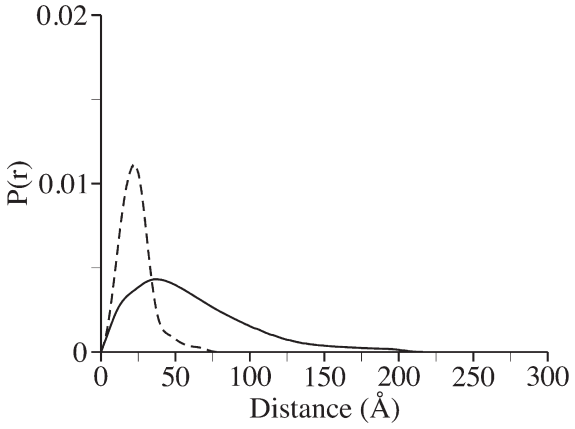


Figure 9. Distance distribution function $P(r)$ extracted from SAXS data using a generalized inverse Fourier transform technique (GIF).[35,36] The ensemble-averaged distance distribution function of osteopontin is shown as a solid line. For comparison, the distance distribution function of a stably folded protein (quail Q83) with a rather compact 3D structure is also shown (---).

solution. Experiments exploiting these features are currently underway in our laboratories.

DISCUSSION

The solution structural features and local dynamics of OPN were investigated by a combined biophysical approach involving meta-structure-based sequence analysis, multidimensional NMR spectroscopy, NMR translational diffusion measurements and ^{15}N relaxation parameters, SRCD spectroscopy, and small-angle X-ray scattering (SAXS). All biophysical data consistently demonstrated that OPN exists as a largely disordered protein in solution and thus constitutes another member of the growing intrinsically disordered protein family. These results are in agreement with a recent report identifying OPN as an intrinsically disordered protein.[13] However, these broad-based and in-depth structural analyses reveal that although OPN is largely unfolded in solution, significant ^{13}C secondary shifts that differ from random coil values indicate that the local secondary structure elements are only transiently formed, lacking the cooperative behavior typical for globular and stably folded proteins. The variation of local mobility along the backbone significantly differs from the profile expected for a random coil protein and is similar to profiles of other intrinsically unfolded or natively unstructured proteins undergoing preferential conformational sampling. In unfolded proteins, residues in the middle of the polypeptide chain are expected to be less mobile compared to terminal segments, leading to a bell-shaped distribution.[61,62] The correlation between deviations from random coil chemical shift values and backbone dynamics provides convincing evidence of the existence of local structure elements involving residues that are located in reasonably well-ordered, hydrogen-bonded, and motionally restricted parts of the polypeptide chain. Local motional restriction implies that residues located in these structurally preformed segments do not populate unfolded (coil-like) conformations but rather exist as cooperatively folded local structural elements (β-strands and α-helices). Interestingly, these preformed and motionally restricted structural motifs comprise the biochemically established protein interaction motifs of OPN.

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

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

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

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

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

■ ASSOCIATED CONTENT

Accession Codes

The SRCD spectrum and associated metadata have been deposited in the Protein Circular Dichroism Data Bank (http://pcddb.cryst.bbk.ac.uk) as accession code CD0001151000.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: robert.konrat@univie.ac.at. Telephone: +43-1-4277-52202. Fax: +43-1-4277-9522.

Funding Sources

This work was supported by Grants P17041 (to K.B.) and P18148 (to M.H.) from the Austrian Science Foundation (FWF) and W 1221-B03, EU-BACRNA, SFB-17, and WWTF LS612 from the FWF (all to R.K.). R.A. is a recipient of a DOC-fFORTE fellowship of the Austrian Academy of Sciences. N.C. is a recipient of a Lise-Meitner FWF fellowship. The SRCD work was supported by project grants from the U.K. Biotechnology and Biological Sciences Research Council (BBSRC) to B.A.W. and by synchrotron beam time grants to B.A.W. Access to the CD1 beamline at ISA is acknowledged under the EU Integrated Infrastructure Initiative (I3), Integrated Activity on Synchrotron and Free Electron Laser Science (IA-SFS), Contract RII3-CT-2004-506008.

■ ACKNOWLEDGMENT

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

■ ABBREVIATIONS

ARC, average residue compactness; CPMG, Carr−Purcell−Meiboom−Gill; CSI, chemical shift index; DOSY, diffusion-ordered spectroscopy; DTT, dithiothreitol; GIF, generalized inverse Fourier transformation; HSQC, heteronuclear single-quantum coherence; IFT, indirect Fourier transformation; IPTG, isopropyl β-D-thiogalactopyranoside; IUP, intrinsically unstructured protein; OPN, osteopontin; MMP, matrix metalloprotease; NOE, nuclear Overhauser effect; NMR, nuclear magnetic resonance; PAGE, polyacrylamide gel electrophoresis; PDDF, pair distance distribution function; PFG, pulsed field gradient; PRE, paramagnetic relaxation enhancement; RDC, residual dipolar coupling; SAXS, small-angle X-ray scattering; SDS, sodium dodecyl sulfate; SIBLING, small integrin-binding ligand N-linked glycoprotein; SRCD, synchrotron radiation circular dichroism.

■ REFERENCES

(1) Bellahcene, A., Castronovo, V., Ogbureke, K. U., Fisher, L. W., and Fedarko, N. S. (2008) Small integrin-binding ligand N-linked

glycoproteins (SIBLINGs): Multifunctional proteins in cancer. *Nat. Rev. Cancer 8* (3), 212–226.

(2) Weber, G. F. (2001) The metastasis gene osteopontin: A candidate target for cancer therapy. *Biochim. Biophys. Acta 1552* (2), 61–85.

(3) Rittling, S. R., and Chambers, A. F. (2004) Role of osteopontin in tumour progression. *Br. J. Cancer 90* (10), 1877–1881.

(4) Wai, P. Y., and Kuo, P. C. (2008) Osteopontin: Regulation in tumor metastasis. *Cancer Metastasis Rev. 27* (1), 103–118.

(5) Hartl, M., Karagiannidis, A. I., and Bister, K. (2006) Cooperative cell transformation by Myc/Mil(Raf) involves induction of AP-1 and activation of genes implicated in cell motility and metastasis. *Oncogene 25* (29), 4043–4055.

(6) Gardner, H. A., Berse, B., and Senger, D. R. (1994) Specific reduction in osteopontin synthesis by antisense RNA inhibits the tumorigenicity of transformed Rat1 fibroblasts. *Oncogene 9* (8), 2321–2326.

(7) Rangaswami, H., Bulbule, A., and Kundu, G. C. (2006) Osteopontin: Role in cell signaling and cancer progression. *Trends Cell Biol. 16* (2), 79–87.

(8) Denhardt, D. T., Noda, M., O'Regan, A. W., Pavlin, D., and Berman, J. S. (2001) Osteopontin as a means to cope with environmental insults: Regulation of inflammation, tissue remodeling, and cell survival *J. Clin. Invest. 107* (9), 1055–1061.

(9) Junaid, A., and Amara, F. M. (2004) Osteopontin: Correlation with interstitial fibrosis in human diabetic kidney and PI3-kinase-mediated enhancement of expression by glucose in human proximal tubular epithelial cells. *Histopathology 44* (2), 136–146.

(10) Okamoto, H. (2007) Osteopontin and cardiovascular system. *Mol. Cell. Biochem. 300* (1−2), 1–7.

(11) Prince, C. W., Oosawa, T., Butler, W. T., Tomana, M., Bhown, A. S., Bhown, M., and Schrohenloher, R. E. (1987) Isolation, characterization, and biosynthesis of a phosphorylated glycoprotein from rat bone. *J. Biol. Chem. 262* (6), 2900–2907.

(12) Schedlbauer, A., Ozdowy, P., Kontaxis, G., Hartl, M., Bister, K., and Konrat, R. (2008) Backbone assignment of osteopontin, a cytokine and cell attachment protein implicated in tumorigenesis. *Biomol. NMR Assignments 2* (1), 29–31.

(13) Yamaguchi, Y., Hanashima, S., Yagi, H., Takahashi, Y., Sasakawa, H., Kurimoto, E., Iguchi, T., Kon, S., Uede, T., and Kato, K. (2010) NMR characterization of intramolecular interaction of osteopontin, an intrinsically disordered protein with cryptic integrin-binding motifs. *Biochem. Biophys. Res. Commun. 393* (3), 487–491.

(14) Konrat, R. (2009) The protein meta-structure: A novel concept for chemical and molecular biology. *Cell. Mol. Life Sci. 66* (22), 3625–3639.

(15) Lees, J. G., Smith, B. R., Wien, F., Miles, A. J., and Wallace, B. A. (2004) CDtool: An integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal. Biochem. 332* (2), 285–289.

(16) Miles, A. J., Wien, F., Lees, J. G., Rodger, A., Janes, R. W., and Wallace, B. A. (2003) Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers. *Spectroscopy 17*, 653–661.

(17) Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res. 32* (Web Server Issue), W668–W673.

(18) Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem. 287* (2), 243–251.

(19) Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry 20* (1), 33–37.

(20) van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem. 191* (1), 110–118.

(21) Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem. 287* (2), 252–260.

(22) Kay, L. E., Saarinen, T., and Keifer, P. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc. 114*, 10663–10665.

(23) Cavanagh, J., Fairbrother, W., Palmer, A. G., and Skeleton, N. (2007) *Protein NMR Spectroscopy*, Academic Press, New York.

(24) Panchal, S. C., Bhavesh, N. S., and Hosur, R. V. (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: Application to unfolded proteins. *J. Biomol. NMR 20* (2), 135–147.

(25) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6* (3), 277–293.

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

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

(28) Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc. 123* (13), 2970–2978.

(29) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33* (19), 5984–6003.

(30) Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D., and Kay, L. E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc. 123* (46), 11341–11352.

(31) Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry 38* (50), 16424–16431.

(32) Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR 2* (6), 661–665.

(33) Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson. 131* (2), 373–378.

(34) Orthaber, D., Bergmann, A., and Glatter, O. (2000) SAXS experiments on absolute scale with Kratky systems unsing water as a secondary standard. *J. Appl. Crystallogr. 33*, 218–225.

(35) Glatter, O. (1977) A new method for the evaluation of small-angle scattering data. *J. Appl. Crystallogr. 10*, 415–421.

(36) Glatter, O. (1979) The interpretation of real-space information from small-angle scattering experiments. *J. Appl. Crystallogr. 12*, 166–175.

(37) Wallace, B. A. (2009) Protein characterisation by synchrotron radiation circular dichroism spectroscopy. *Q. Rev. Biophys. 42* (4), 317–370.

(38) Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR 13* (3), 289–302.

(39) He, B., Mirza, M., and Weber, G. F. (2006) An osteopontin splice variant induces anchorage independence in human breast cancer cells. *Oncogene 25* (15), 2192–2202.

(40) Kazanecki, C. C., Uzwiak, D. J., and Denhardt, D. T. (2007) Control of osteopontin signaling and function by post-translational phosphorylation and protein folding. *J. Cell. Biochem. 102* (4), 912–924.

(41) Boskey, A. L., Maresca, M., Ullrich, W., Doty, S. B., Butler, W. T., and Prince, C. W. (1993) Osteopontin-hydroxyapatite interactions in vitro: Inhibition of hydroxyapatite formation and growth in a gelatin-gel. *Bone Miner. 22* (2), 147–159.

(42) Goldberg, H. A., and Hunter, G. K. (1995) The inhibitory activity of osteopontin on hydroxyapatite formation in vitro. *Ann. N.Y. Acad. Sci. 760*, 305–308.

(43) Hunter, G. K., Kyle, C. L., and Goldberg, H. A. (1994) Modulation of crystal formation by bone phosphoproteins: Structural specificity of the osteopontin-mediated inhibition of hydroxyapatite formation. *Biochem. J. 300* (Part 3), 723–728.

(44) Worcester, E. M., and Beshensky, A. M. (1995) Osteopontin inhibits nucleation of calcium oxalate crystals. *Ann. N.Y. Acad. Sci. 760*, 375–377.

(45) Gericke, A., Qin, C., Spevak, L., Fujimoto, Y., Butler, W. T., Sorensen, E. S., and Boskey, A. L. (2005) Importance of phosphorylation for osteopontin regulation of biomineralization. *Calcif. Tissue Int. 77* (1), 45–54.

(46) Mayer, O., Rajkowitsch, L., Lorenz, C., Konrat, R., and Schroeder, R. (2007) RNA chaperone activity and RNA-binding properties of the *E. coli* protein StpA. *Nucleic Acids Res. 35* (4), 1257–1269.

(47) Linding, R., Russell, R. B., Neduva, V., and Gibson, T. J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res. 31* (13), 3701–3708.

(48) Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins 42* (1), 38–48.

(49) Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Inf. Ser. 10*, 30–40.

(50) Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics 21* (16), 3433–3434.

(51) Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) Protein disorder prediction: Implications for structural proteomics. *Structure 11* (11), 1453–1459.

(52) Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol. 337* (3), 635–645.

(53) Zeev-Ben-Mordehai, T., Rydberg, E. H., Solomon, A., Toker, L., Auld, V. J., Silman, I., Botti, S., and Sussman, J. L. (2003) The intracellular domain of the *Drosophila* cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded. *Proteins 53* (3), 758–767.

(54) Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science 293* (5529), 487–489.

(55) Bernado, P., Bertoncini, C. W., Griesinger, C., Zweckstetter, M., and Blackledge, M. (2005) Defining long-range order and local disorder in native α-synuclein using residual dipolar couplings. *J. Am. Chem. Soc. 127* (51), 17968–17969.

(56) Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol. 268* (1), 170–184.

(57) Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W., and Blackledge, M. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc. Natl. Acad. Sci. U.S.A. 102* (47), 17002–17007.

(58) Bertoncini, C. W., Jung, Y. S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M., and Zweckstetter, M. (2005) Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein. *Proc. Natl. Acad. Sci. U.S.A. 102* (5), 1430–1435.

(59) Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2005) Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc. 127* (2), 476–477.

(60) Hartl, M., Matt, T., Schuler, W., Siemeister, G., Kontaxis, G., Kloiber, K., Konrat, R., and Bister, K. (2003) Cell transformation by the v-myc oncogene abrogates c-Myc/Max-mediated suppression of a C/EBP *β*-dependent lipocalin gene. *J. Mol. Biol. 333* (1), 33–46.

(61) Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1997) Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry 36* (9), 2390–2402.

(62) Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR 6* (2), 153–162.

(63) Tamiola, K., Acar, B., and Mulder, F. A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc. 132* (51), 18000–18003.