Polyelectrolyte Domains and Intrinsic Disorder within the Prismatic Asprich Protein Family[†]

Katya Delak, Sebastiano Collino, and John Spencer Evans*

Laboratory for Chemical Physics, New York University, 345 East 24th Street, New York, New York 10010

Received January 24, 2009; Revised Manuscript Received March 2, 2009

ABSTRACT: A number of fully functional proteins have been identified to exist in a partially or fully disordered state. These intrinsically disordered proteins (IDP) are recognized as an important sequence class that fulfill many roles. A number of biomineral-associated proteins, particularly those which possess polyelectrolyte domains, represent potential members of the IDP class. This report describes a bioinformatics study of a ten member polyanionic sequence biomineralization protein family, Asprich, and the experimental characterization of the conserved N- and C-terminal regions found within seven members of this family. Using protein disorder prediction algorithms (DPROT, PONDR, GLOBPLOT), we confirm that all ten Asprich protein sequences are disordered, and that two polyelectrolyte domains within each protein contribute to the disorder scoring. Using synthetic peptides which model the conserved N- (F1,48 AA) and C-terminal (F2, 42 AA) domains, we determine that both domains are globally disordered and remain so in the presence of Ca(II). However, F1 and F2 possess differing proportions of extended beta strand relative to random coil structure and sequence spacing of Asp, Glu residues. As a result, the F2 sequence possesses a higher anionic surface charge density, solvent accessibility, and greater degree of local conformational response to Ca(II). These differences may explain why F1 and F2 differ with regard to step growth kinetics, mineral modulation, and metal ion complexation, and possibly distinguish the molecular role(s) that each domain conveys to the Asprich protein family. Structural and surface charge density features may also control the function of Asp, Glu polyelectrolyte domains within other IDP proteins.

There are proteins which are fully functional yet exist in either a partially or fully unfolded state (*1*−*3*). These proteins have been named intrinsically disordered proteins or IDPs[1] (*2*). Disordered sequences are dynamic, and their global structures can be partially or completely extended as a result of a combination of factors, including low sequence complexity, compositional bias toward aromatic and hydrophilic residues, and high flexibility (*2, 3*). The amino acid composition of IDPs are less complex and are dominated by disorder-promoting residues (E, K, R, G, Q, S, P, A) (*2*). Recent studies also indicate that polyelectrolyte sequence regions (i.e., polyanionic, polycationic, or heterogeneously charged) typically exist in a disordered state (*2*). With regard to molecular recognition, at least two advantages are offered by disordered sequences. First, these sequences can adopt multiple metastable conformations, which, in turn, allows recognition of several targets with high specificity and low affinity (i.e., binding promiscuity) (*1, 2*). Second, the binding of the unfolded protein sequence to its target(s) often results in a disorder-to-order conformational transition within the unfolded protein, which would decrease conformational entropy and make highly specific interactions possible (*2*). Thus, IDPs represent an enigmatic yet important set of proteins with unusual structural and functional traits that distinguish them from folded proteins.

Recently, studies of the biomineralization proteins starmaker (*4*), amelogenin (*5*), and AP7 (*6*) established that these sequences are either partially or fully disordered and belong to the IDP class. This finding raises the possibility that there may be other biomineralization proteins that are IDPs as well, but have yet to be identified as such. Given the inherent disorder in many polyelectrolyte domains (*2*), it stands to reason that those biomineralization proteins that possess polyelectrolyte domains (*7*−*9*) may also exhibit disorder within these domains. Conceivably, the multiply charged regions of these proteins could bind to multiple "targets" (i.e., binding promiscuity) (*2*) such as metal ions, other proteins, or mineral surfaces (*4*−*9*) during the mineralization processes and undergo disorder-to-order transformations during binding. Hence, disordered polyelectrolyte domains

[†] This work was supported by funding from the National Science Foundation Division of Materials Research/BMAT (DMR-0704148 to J.S.E.). The INOVA 600 NMR spectrometer was supported by a Defense University Research Instrumentation Program award from the Army Research Office (W911NF-04-1-0214), and this paper represents contribution number 48 from the Laboratory for Chemical Physics, New York University.

* To whom correspondence should be addressed. E-mail: jse1@nyu.edu. Tel: 212-998-9605. Fax: 212-995-4087.

[1] Abbreviations: IDP, intrinsically disordered or unfolded protein; MS/TOF, mass spectrometry time-of-flight; PPII, polyproline type II; F1, fragment 1, 48 AA sequence representing a portion of the conserved N-terminus of Asprich proteins "a" through "g"; F2, fragment 2, 42 AA sequence representing the conserved C-terminus of Asprich proteins "a" through "g"; SA/MD, simulated annealing molecular dynamics; DPROT, Disordered Protein Predictor; PONDR, Predictor of Naturally Disordered Regions; RC, random coil; ACC, amorphous calcium carbonate.

10.1021/bi900113v CCC: $40.75
Published on Web 04/03/2009

and their binding promiscuity (*2*) could fulfill a number of roles in protein-mediated biomineralization processes.

An example of a polyelectrolytic sequence-containing family of proteins is Asprich, a group of ten proteins found in the prismatic (calcite) shell layer of the mollusk, *Atrina rigida* (*9*). This protein family is linked to the stabilization of amorphous calcium carbonate (ACC) and eventual crystallization of ordered calcite arrays in the prismatic layer (*10*). Although the Asprich protein family is not yet available in sufficient quantities for biophysical studies, recent studies of synthetic peptides representing the highly conserved N- and C-terminal sequences of seven Asprich proteins ("a" through "g") (*8*) reveal some important features. These terminal sequences modulate calcite morphology, form amorphous deposits on crystal surfaces, and stabilize both ACC and vaterite in vitro (*8*). Hence, these conserved Asprich terminal sequences are functionally active and perform tasks that are similar to those of the parent protein family. However, we do not yet possess a detailed understanding of the structure of these conserved terminal domains, or how this structure influences domain functionality (*8, 10*).

To broaden our understanding of polyelectrolyte sequences and IDP proteins, we initiated a combined bioinformatics/biophysical study of the Asprich protein family. Using three disorder predictive algorithms [DPROT (*11*), GLOBPLOT2.3 (*12*), PONDR (*13*)], we confirm that all ten members of the Asprich family score as IDPs and we have identified key regions of disorder within each sequence. To verify these predictive studies, we created synthetic peptides representing the highly conserved N- (fragment 1 or F1, 48 AA) and C- (fragment 2 or F2, 42 AA) terminal sequences that are found in the Asprich "a" through "g" protein series (*8*). Experimentally, we confirm that both F1 and F2 bind multiple metal ions and are globally disordered in the metal bound state. What distinguishes F1 from F2 are the differences in linear Asp, Glu spacing and the proportion of extended beta strand relative to random coil segments. These elements create a higher anionic surface charge density and solvent accessibility in F2 relative to F1. This, in turn, endows F2 with a higher metal ion complexation capacity and a more significant local conformational response to Ca(II) ion binding. We believe that these molecular differences are responsible for the in vitro functions of each conserved domain as separate entities and as participants in Asprich-directed ACC stabilization and/or prismatic crystal development.

MATERIALS AND METHODS

Prediction of Disordered Regions within the Asprich Family of Sequences. The ten Apsrich protein sequences, "a" through "g" (accession numbers AAU04812, AAU04809, AAU04806, AAU04810, AAU04808, AAU04807, and AAU04811, respectively) and "1" through "3" (accession numbers AAU04814, AAU04813, and AAU04815, respectively) were obtained from the National Center for Biotechnology Information. Asprich sequences were analyzed using three different disorder prediction algorithms. The first was DPROT (526 ordered, 82 disordered protein data set), which is a position-specific scoring matrix that utilizes evolutionary information of amino acid sequences to predict the existence of disordered proteins (threshold = 0) (*11*). The second algorithm is GlobPlot (version 2.3) (*12*), which utilizes propensities to identify whether or not a given amino acid in a sequence is either random coil (RC) or a defined secondary structure. The third algorithm is PONDR (VL-XT predictor) (*13*), which relies on the fractional composition of particular amino acids or hydropathy to determine the percentage of disorder within a given primary sequence. The scores obtained from these algorithms were graphically compared to one another to determine the presence and extent of disorder in each Asprich sequence (Figure 1). In the case of GlobPlot2.3, we utilized this algorithm to identify specific disordered sequence regions within each Asprich protein (Figure 1).

Asprich Fragment Peptide Synthesis and Sample Preparation. The free N-terminal, C^{α}-amide-capped 48-AA fragment 1 (F1) and the N-acetyl-capped, free alpha-carboxyl 42 AA fragment 2 (F2) polypeptides were synthesized and purified as previously described (*8*). C- and N-terminal alpha amide capping procedures were performed to simulate peptide bond attachment (*8*). The primary sequences of each fragment are presented in Figure 7. Stock solutions of F1 and F2 were created by dissolving each sequence in unbuffered deionized distilled water (UDDW) to minimize aggregation (*15*) and adjusted to a final pH of 7.2 using microliter amounts of 1 N NaOH.

Ion Trap Mass Spectrometry Metal Ion Binding Studies. Subsequent dilutions of peptide stock solutions were accomplished using UDDW to create final peptide concentrations of 8 μM for ion trap MS/TOF experiments. For metal-binding experiments, the appropriate amounts of metal salt stock solutions ($CaCl_2$, $CdCl_2$, $LaCl_3$, $EuCl_3$, each 5 M in UDDW, 99.99% pure, Sigma-Aldrich Chemicals) were added to each peptide sample to create final peptide:metal ion ratios of 1:10 (*16*). Ion trap experiments were conducted on an Agilent LC/MSD 1100 ion trap mass spectrometer with an electrospray ionization source. Samples were injected using a 250 μL syringe loaded into a stepper motor that delivered sample to the instrument at a rate of 10 μL/min. The mass spectrometer was operated using UltraScan mode with an N_2 nebulizing gas pressure of 15 psi with an N_2 drying gas flow of 4 L/min at 350 °C. Positive ionization mode was employed with a skimmer voltage of −40 V and a capillary exit voltage of −140 V, resulting in a capillary potential of 100 V.

Solution State NMR Spectroscopy. Apo F1 and F2 samples were each prepared at a concentration of 450 μM in 90% v/v UDDW, 10% v/v D_2O (99.9% atom D, Cambridge Isotope Laboratories), and the sample pH adjusted to 7.2 using microliter volumes of NaOH with the peptide acting as its own buffer. All NMR experiments were conducted at 293 K on a narrow bore Varian INOVA 600 z-PFG Digital NMR spectrometer outfitted with a triple-channel HCN 5 mm z-PFG Nalorac probe. Pulsed field gradient (PFG) total correlation experiments (TOCSY; t_{mix} = 60, 70, 80 ms) (*16, 17*) were used to identify amino acid proton resonances. Spectral data from the t_{mix} = 80 ms TOCSY experiment were used for the determination of $^3J_{HNH\alpha}$ values from peak linewidths (*18*). PFG NOESY experiments (t_{mix} = 50, 100, 150 and 200 ms) were used for sequential assignments and nOe quantitation (*16, 17*). To determine the effects of Ca(II) ions on the global and local conformational preferences of F1 and F2, parallel PFG NOESY experiments were carried out in the presence of Ca(II) with a Ca(II):

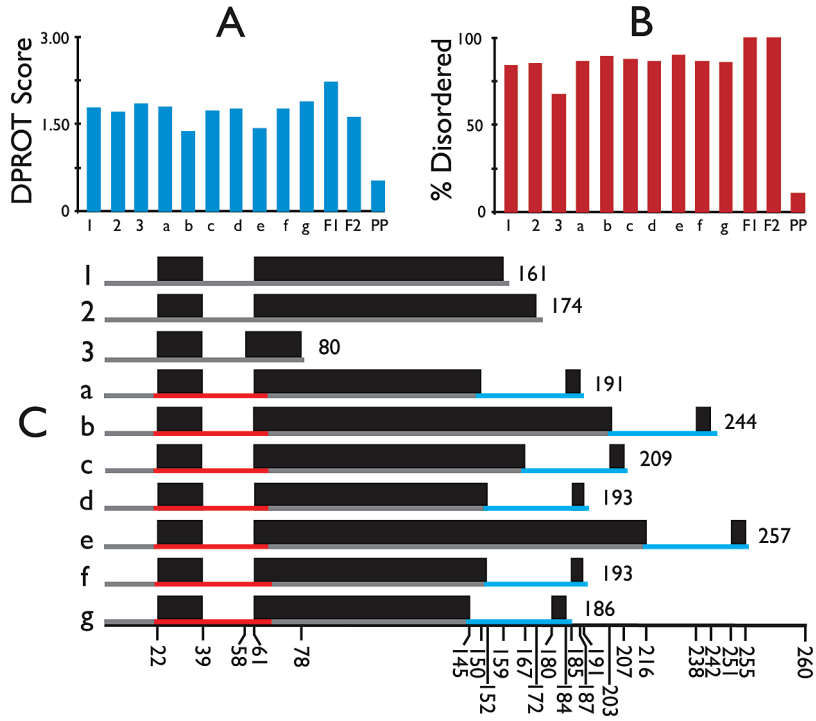

FIGURE 1: Algorithm prediction of disorder within Asprich sequences. (A) DISPROT score (total sequence). (B) PONDR percent disorder (total sequence). In (A) and (B), the ten Asprich proteins are listed along the *x*-axis by their single letter code; "F1" and "F2" refer to individual 48 AA fragment 1, 42 AA fragment 2 sequences, and "PP" refers to the 19 AA prepro leader sequence found at the N-terminal end of each Asprich protein. (C) GLOBPLOT2.3 sequence score. Here, disordered sequence regions are presented as black horizontal bars at designated locations within the primary sequence (*x*-axis). The total sequence length is portrayed by the gray line and is numerically represented at the end of each sequence. The location of the F1 (red) and F2 (blue) domains are denoted on the gray sequence line. The start and stop of each disordered region is numerically identified on the *x*-axis.

peptide ratio of 10:1, identical to that utilized in our ion trap/MS experiments (see previous section). Water suppression was achieved in all experiments using 3-9-19 *z*-axis WATERGATE PFG schemes, with NOESY experiments employing an additional soft flip-back pulse prior to the mixing time (*16, 17*). All NMR experiments were processed using NMRPipe software (NIH) and visualized using Sparky software (SPARKY 3, University of California, San Francisco). NMR acquisition and processing parameters are provided in detail with the corresponding figure legends.

Structure Refinement: Simulated Annealing Molecular Dynamics. The initial peptide PDB files were generated using MacPyMol (DeLano Scientific, LLC, Palo Alto, CA). Torsion angles of the initial structures were not deliberately specified. Peptides were constructed assuming a neutral pH scenario in which all aspartate and glutamate side chains were negatively charged, side chains on lysine and asparagine were positively charged, the N-terminus was positively charged and the C-terminus was negatively charged. These structures differed from those used in the experimental protocols in that the capping moieties were not explicitly represented. All calculations were carried out using the XPLOR software package (version 3.1). Protein parameters were defined using the default XPLOR protein parameter file, protein.par, which is based on the CHARMM force field (*19*). NMR interproton distance restraints based on experimental NOESY data were categorized as strong (1.8−2.5 Å), medium (2.5−3.5 Å) or weak (3.5−5.0 Å), and these were calibrated using alanine Hα-Hβ NOESY cross peaks with a reference distance of 2.4 Å (*16, 17, 20*). Pseudoatom corrections were applied as necessary and used as part of the restraints input file (*21*). Coupling constant restraints were excluded from the calculations. A total of 94 and 67 nOe restraints were specified for F1 and F2, respectively. The NOE restraining function was defined as a soft-square well potential with a force constant of 50 kcal mol^{-1} $Å^{-2}$. A 4.5 Å cutoff for nonbonding interactions was used. Implicit solvation was achieved through the use of a distance-dependent dielectric of 78.5 (*16, 17, 19*).

The simulated annealing experiments were performed using a modified version of the XPLOR sa.inp file wherein each structure's coordinates are randomized by a chaotic variation of each phi and psi angle (*16, 17*). This is followed by 30 ps of equilibrium dynamics at 300 K that were added after the initial minimization and before heating the structure to 2000 K. The system is cooled from 2000 to 300 K at 100 K/step. The resultant structures are then energy minimized with the Powell method. Calculations are carried out with a Verlet algorithm integrator in 1 fs time steps. Temperature coupling was achieved with a Langevin-type dynamics with zero random forces and a scaled friction coefficients. Each SA/MD run generated a pool of 1000 structures in the initial ensemble, with 100 structures chosen from this pool based upon the fewest nOe and ϕ angle violations. From this 100 member subset, 10 lowest energy structures were selected to form the conformer library for further analyses.

Lowest energy coordinate files were visualized using the Visual Molecular Dynamics (VMD, v 1.8.6) and Python Molecular Viewer (PMV, v 1.5.2) (*22*) software packages. Solvation energies (vacuum to solvent transfer) and molecular surface plots were created for lowest energy conformers with the Adaptive Poisson−Boltzmann Solver (APBS) solvation calculation module (*22*) within the PMV package. Solvation parameters include a linearized Poisson−Boltzmann

method using single Debye−Huckel boundary conditions and a spline-based surface smoothing method, a protein dielectric of 2.0, a solvent dielectric of 78.54, and grid dimensions of 65 Å × 65 Å × 65 Å. The grid spacings for F1 and F2 were 0.821 × 0.802 × 0.883 Å and 1.431 × 0.914 × 0.667 Å, respectively. A solvent radius of 1.4 Å, a system temperature of 298 K, and a system ionic strength of 0.01 M were employed (the latter was chosen to mimic the low ionic strength conditions of our NMR samples *16, 17*).

RESULTS

Disordered Sequence Prediction of the Asprich Family. Recently, bioinformatics algorithms have been developed and applied to determine the presence of disordered regions within protein sequences (*3, 11*−*13*). We applied two of these algorithms, DPROT (*11*)(Figure 1A) and PONDR (*13*) (Figure 1B), to determine overall percent disorder within each Asprich protein sequence ("1" through "3", "a" through "g"), and for comparison, for the individual, highly conserved F1, F2, and the prepro leader sequences (MKGLAILIAIAAL-LAVSHP) found within Asprich "a" through "g" proteins (*9*). As shown in Figure 1, DPROT and PONDR score all ten Asprich proteins as "disordered", meaning that the obtained scores exceed a threshold value for disorder proteins [for DPROT, this threshold value is 0.5 (*11*); for PONDR, >50% (*13*) disorder). The variation in disorder scoring (10−20%) is fairly consistent across the Asprich sequences, with Asprich "3" and "b" exhibiting the lowest PONDR and DPROT scores, respectively. For comparison, the 19 AA hydrophobic prepro leader sequence scores very low on the disorder scale and the two terminal conserved sequences, F1 and F2, score very high on the same scales.

Using GlobPlot2.3 (*12*) we identified two common disordered regions within the primary sequences of all ten Asprich proteins (Figure 1C). These are as follows: (a) Sequence regions 22−39. This corresponds to the −VFKRSLS-DPSDDGGANDV− sequence stretch of F1 (Figure 7) and includes a 5 AA portion of the conserved 7 AA −KPVFKRS− "cationic" domain of Asprich "a" through "g" (*9*). (b) A variable length polyanionic sequence stretch which initiates at residue 58 in Asprich "3" and at residue 61 in all other Asprich proteins. As one can observe from Figure 1C, this variable length anionic region comprises the longest disordered sequence stretch in all ten Asprich proteins and most likely accounts for the overall predicted disorder of each protein (Figure 1A,B). Note also that this variable length polyanionic region also encompasses the sequence fragment −NDVDDEE− of F1.

In addition to these two common regions, a third region of disorder was identified by GlobPlot2.3 to exist in the Asprich "a" through "g" series of proteins. This is a conserved 5 AA sequence is located near the C-terminal region and corresponds to the −EDVAD− sequence found in the F2 sequence (Figure 7). Interestingly, this disordered 5 AA segment is not present in the Asprich "1" through "3" series (Figure 1). The remaining Asprich sequence regions were not identified as disordered by GlobPlot2.3 (Figure 1; note regions in gray). Presumably, some degree of ordering exists within these sequence blocks. From these analyses, we conclude that all ten Asprich proteins are IDPs, primarily due to the presence of a large, internal polyanionic segment that comprises all members of this family.

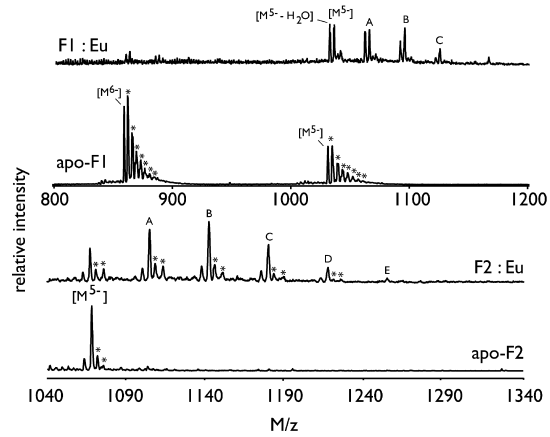

FIGURE 2: Representative ion trap/TOF mass spectrometry of Asprich terminal sequences in the apo state and in the presence of 10:1 $EuCl_3$. F1 = fragment 1, F2 = fragment 2. The full set of peptide:metal ion adducts for both sequences in the presence of Cd(II), La(III), and Eu(III) is provided in Figures S1 and S2 and Tables S1 and S2 in the Supporting Information. We also note that both fragments form multiple sodium ion adducts (denoted by "*"), most likely arising from residual buffer salts from HPLC purification and/or subsequent NaOH sample titration. For F1, peptide:Eu(III) adducts are as follows: A = $[M^{5-} + Eu^{3+} - 6H^+]^{8-}$; B = $[M^{5-} + 2(Eu^{3+}) - 9H^+]^{8-}$; C = $[M^{5-} + 3(Eu^{3+}) - 12H^+]^{8-}$. For F2, peptide:Eu(III) adducts are as follows: A = $[M^{3-} + Eu^{3+} - 2H^+]^{2-}$; B = $[M^{3-} + 2(Eu^{3+}) - 5H^+]^{2-}$; C = $[M^{3-} + 3(Eu^{3+}) - 8H^+]^{2-}$; D = $[M^{3-} + 4(Eu^{3+}) - 11H^+]^{2-}$; E = $[M^{3-} + 5(Eu^{3+}) - 14H^+]^{2-}$. For more information on F1 and F2 adducts formed in the presence of Cd(II) and La(III), please consult Figures S1 and S2 and Tables S1 and S2 in the Supporting Information.

Metal Ion Sequestration and the Effects of Metalation on Conserved Asprich N- and C-Terminal Sequence Folding. Recent studies indicate that protein−metal ion complexation (i.e., metalation) can induce folding or self-association or promote protein−protein interactions in intrinsically disordered proteins (i.e., metalation) (*14, 23*). In some cases, the metalation effect is cation-specific (*23*). In the case of Asprich, we acknowledge that Ca(II) ions in the prismatic layer could serve as a putative folding agent for these proteins, which, in turn, might trigger ACC synthesis, stabilization, or other mineralization events (*10*). With this in mind, our next step was to first assess the metal ion binding or sequestration capabilities of disordered sequences common to most of the Asprich family, then determine if Ca(II) ion metalation leads to a disorder-to-order transformation.

Using F1 and F2 synthetic peptides as models for the highly conserved N- and C-terminal sequences of Asprich "a" through "g" (*8, 9*) we assessed the metalation capabilities of each sequence using ion trap MS-TOF/metal ion-induced mass shift experiments and Ca(II) metal ion analogues, Cd(II), La(III), and Eu(III) (Figure 2; Figures S1 and S2 and Tables S1 and S2 in the Supporting Information) (*6, 16, 17*). As a result of the high Asp, Glu content in both peptides, we observe multiple peptide:metal ion adduct species in the presence of all Ca(II) analogue metal ions [i.e., Cd(II), La(III), Eu(III)] (*6, 16, 17*). Specifically, F1 forms 1:1, 1:2, and 1:3 peptide:metal ion adducts with Cd(II), La(III), and Eu(III); and F2 forms 1:1 and 1:2 adducts with Cd(II) and 1:1, 1:2, 1:3, 1:4, and 1:5 adducts with La(III) and Eu(III). At this time we do not fully understand why stoichiometric differences exist for Cd(II) versus Eu(III) and La(III). However, what is clear is that both F1 and F2 sequences are quite capable of forming stable complexes with Ca(II) ion

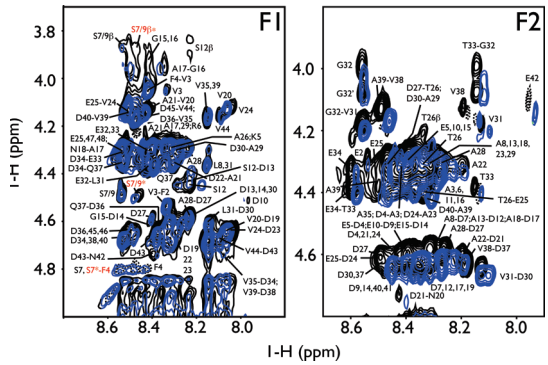

FIGURE 3: ^{1}H NMR PFG-NOESY (t_{mix} = 200 ms, 293 K) overlay spectra (NH−CH_α fingerprint region) of F1 (A) and F2 (B) polypeptides in the apo state (black) and in the presence of Ca(II) (peptide:Ca = 1:10, blue, same stoichiometry as per ion trap/MS experiments), pH 7.2. Identifiers in red indicate residues that are in conformational exchange. Spectra were acquired using identical acquisition parameters, sample volumes and peptide concentrations. Acquisition parameters include a relaxation delay of 1.3 s, a spectral window of 11 ppm in the ω_2 dimension and 10 ppm in ω_1 dimension, ^{1}H 90° pulse = 5.0 μs, and 64 transients per experiment, with 2048 and 256 complex points acquired in t_2 and t_1 respectively. The carrier frequency is centered on the water resonance. Parameters for the *Z*-gradients G0 (homospoil) and G3 (spin−echo) were identical at 10.64 G/cm with a duration of 3 ms and a stabilization time of 3 ms. The hypercomplex States-TPPI phase sensitive method was used to process both spectra, with zero-filling in both dimensions. Solvent filtering was applied in the time domain, and cosine bell squared window functions were applied to the FIDs prior to Fourier transform. Proton chemical shifts are referenced to internal d_4-TSP.

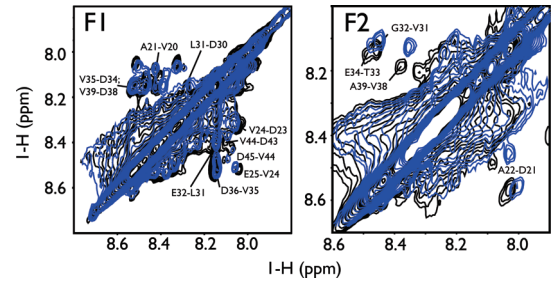

FIGURE 4: ^{1}H NMR PFG-NOESY (t_{mix} = 200 ms, 293 K) overlay spectra (NH−NH fingerprint region) of F1 and F2 polypeptides in the apo state (black) and in the presence of Ca(II) (peptide:Ca = 1:10, blue), pH 7.2. Acquisition and processing parameters described in Figure 3 are identical to those employed here.

analogues, with F2 forming higher-order adduct complexes compared to F1.

As a follow up to our mass spectrometry experiments, Ca(II) induced global folding propensities of both F1 and F2 were assayed using circular dichroism (CD) spectrometry (Figure S3 in the Supporting Information). Here, we find that Ca(II) metalation does not promote global folding, disorder-to-order transitions, or detectable alterations in peptide secondary structure in either sequence. Presumably, this absence of global folding in the presence of Ca(II) ions may be important to the function of these conserved sequences within the supersaturated environment of the mollusk prismatic layer (*7, 8, 16*). However, as we shall see below, Ca(II) interactions do lead to local conformational changes within F1 and F2.

Solution Structures of Polyelectrolyte Domains. To more precisely determine the extent of disorder and location of residual secondary structure within the conserved Asprich N- and C-terminal sequences, we compiled solution NMR data sets for F1 and F2 (Figures 3, 4, and 5 and Tables S3 and S4 in the Supporting Information) (*24*) and quantitated nOe's and ^3J coupling constants to identify the location of random and nonrandom coil structure in each sequence.

For F1, we note that the ^3J coupling constants fall into one of two categories. The first is $^3J > 8$ Hz, which are representative of extended beta strand structure (*20, 25, 26*) and comprise 28% of the sequence. The second category is $^3J = 6$ to 8 Hz, which are representative of random coil structure (*20, 25, 26*) and comprise 25% of the sequence (Figure 5). Only one residue, E41, possesses a $^3J < 6$ Hz which is characteristic of α-helical structure (*20, 25, 26*). These findings are also corroborated by NOESY data, which reveal the presence of a limited set of intraresidue $d_{alphaN(i,i)}$ and sequential $d_{alphaN(i,i+1)}$, $d_{NN(i,i+1)}$ backbone NOE's, only a few backbone−side chain NOE's, and no detectable side chain−side chain d_{SC-SC} NOE's (Figures 3, 4, and 5) (*1, 25, 27*). We find that there are six short regions (V3−F4, V20−A21, V24−E25, L31−E32, D34−V35, and D38−D40) that are extended beta strand as evidenced by their large, nonrandom coil nOe ratios (*25, 27*). We note that large NOE ratios were also obtained for the disordered, polyproline type II (PPII) PEVK titin domains (*25*), and it is likely that these extended beta segments give rise to the weak PPII-like CD ellipticities observed for F1 (*8c*) and other polyelectrolyte sequences (*28*). Collectively, these conformational features describe a polypeptide sequence that is unfolded and extended, possessing alternating segments of RC and extended beta strand (Figure 5). Given the highly anionic composition of F1, this mostly extended structure would help to minimize anionic side chain−side chain charge repulsion by spacing Asp and Glu side chains as far away from one another as possible (*28*).

An unique feature of F1 is the presence of conformational exchange (Figure 3), as evidenced by the presence of duplicated NOESY NH−CH_α cross peaks for residues F4, S7 and S9. Duplication suggests the presence of two conformations, presumably in slow exchange with one another (*29*), and volume integrations of the duplicated cross peaks correspond to a 2:1 ratio of one conformational state relative to the other. Coincidentally, these three residues lie downstream of an isolated Pro residue (P11) and exist within a cationic region, KPVFKRS, that is adjacent to a negatively charged sequence segment, −LSDPSDDGGA− (Figure 7). As reported elsewhere, conformational exchange can result either from cis−trans proline isomerization within proteins and polypeptides or from hingelike motion originating from the Pro imido ring (*16, 29, 30*). Either phenomenon can give rise to duplicated NMR cross peaks (*29, 30*). Hence, the presence of conformational exchange within the N-terminal segment of F1 is most likely in response to the motional behavior of P11, and may be influenced by electrostatic interactions between anionic and cationic side chains near the N-terminus.

In contrast to F1, F2 exhibits fewer sequential backbone nOe's, only one backbone−side chain nOe, and nOe ratios that are more random coil in nature (Figures 3, 4, and 5). Note that sequence blocks A3−D4, A23−D24, T33−E34, and V38−A39 possess backbone nOe ratios that exceed the random coil threshold (*25, 27*). From this, we conclude that

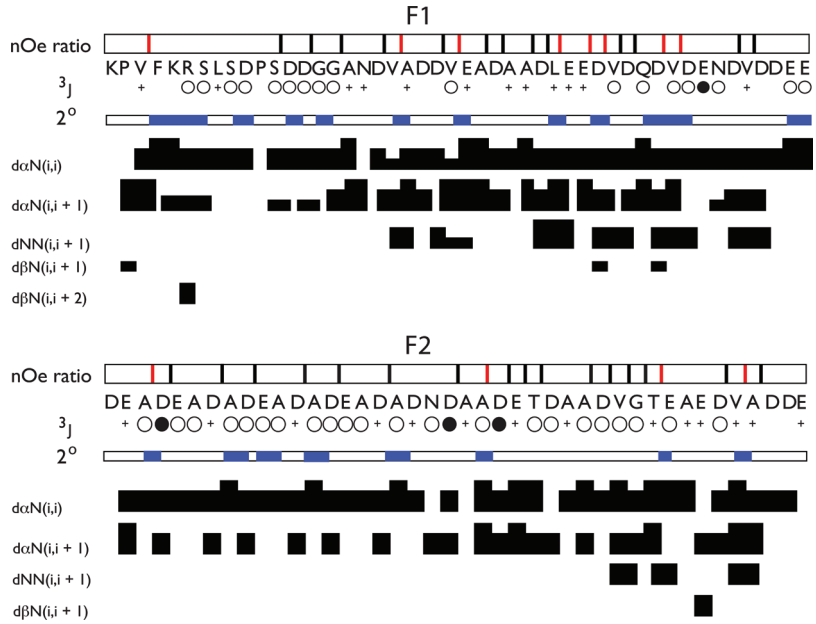

FIGURE 5: Summary of sequence-specific NMR parameters for apo-F1 and F2 at pH 7.2, 293 K. This figure summarizes $^3J_{HNH\alpha}$ coupling constants (^3J) and intraresidue, interresidue, sequential, and medium-range backbone NOE's, and NOE ratios. The schematic representation of coupling constant values immediately below the amino acid sequence are designated as follows: filled circles identify residues with $^3J < 6$ Hz, indicative of local alpha helical-type conformation; open circles correspond to $^3J > 8.0$ Hz, indicative of residues in extended chain conformation; (+) symbols identify residues with $^3J = 6.0-8.0$ Hz which corresponds to other conformations, including random coil (RC). For the proton−proton NOE connectivities, the three different bar heights represent strong, medium, and weak NOE intensities. No long-range NOE intensities were observed. "nOe ratio" refers to backbone NOE intensity ratios [αN(i, i + 1)/αN(i, i), abbreviated as αN/αN] and/or [αN(i, i + 1)/NN(i, i + 1), abbreviated as αN/NN] (*24, 27*) that were calculated for observable, resolvable nOe cross peaks and are denoted as color-coded bars along the sequence number line (black = random coil, αN/NN ≤ 1.4; αN/αN ≤ 2.3; red = extended beta strand, αN/NN > 1.4; αN/αN > 2.3). The absence of a bar indicates that an NOE ratio could not be obtainable for this residue. 2° indicates the tentative secondary structure assignment based upon NOE and ^3J data in this figure; blue = extended beta strand, and white = random coil or unstructured. The terminology and presentation of this figure follows the standardization set forth in ref *24* with some modifications.

these nonrandom coil regions are most likely extended beta strand in nature (*25, 27*). This is also confirmed by ^3J values, where a nearly 2:1 preference for extended beta strand versus random coil is observed ($^3J > 8$ Hz, 22 residues, versus $^3J = 6 - 8$ Hz, 14 residues), and three residues are identified to be alpha helical ($^3J < 6$ Hz). Hence, F2 is unfolded and disordered, possessing a larger proportion of RC regions relative to extended beta strand (Figure 5). This conformational arrangement is the opposite of what we find for F1. As per F1, an unfolded conformation would be advantageous for minimizing Asp, Glu side chain−side chain charge repulsion within the F2 sequence (*28*).

Evidence for Ca(II)-Induced Local Perturbations within Asprich Polyelectrolyte Peptide:Ca(II) Adducts. Although our CD data indicates that Ca(II) ions do not induce folding within either F1 or F2 (Figure S3 in the Supporting Information), it is known that polyanionic biomineralization protein sequences do undergo local structural alterations and changes in backbone motion and dynamics in the presence of Ca(II) (*7b, 31*). To determine if Ca(II) interactions affect local regions within F1 and F2, we performed PFG NOESY spectra overlay comparisons of both sequences in the apo form and in the presence of excess Ca(II) (10:1 Ca(II):peptide) (Figures 3 and 4) that mirrors the stoichiometry level utilized in our ion trap experiments (Figure 2). We observe that both F1 and F2 exhibit Ca(II)-induced alterations in backbone NOE cross peak intensities and/or ^{1}H resonance frequencies for Asp, Glu residues and the neighboring amino acids which flank these sites. However, the local conformational response in F2 is more dramatic, as evidenced by ^{1}H resonance frequency shifts up to 0.1 ppm and significantly attenuated NOE cross peak intensities (Figures 3, 4). From this, we conclude that Ca(II) induced local perturbations in conformation and motion occur within both sequences, with F2 > F1 in terms of conformational and motional/dynamic response to Ca(II)−peptide binding.

Model Building. To gain insights into the possible molecular configurations of conserved N- and C-termini of the Asprich "a" through "g" series of proteins, we utilized the F1 and F2 NOE data set as restraints for molecular dynamics simulated annealing using XPLOR-NIH. As seen in past studies (*5, 15−17*), conformationally disordered proteins typically possess a limited number of NMR backbone restraints and this is certainly true for F1 and F2 (Figure 5). Thus, we advise that the conformations generated by our simulations should be considered as qualitative in nature and we expect that more substantial modeling attempts will be forthcoming.

A comparison of the backbone aligned structural ensembles ($n = 10$ for each sequence) that fit the target NMR data set shows overall agreement in terms of global, extended conformation with some evidence of coil-like structure (Figure S4 in the Supporting Information). The root-mean-square deviation (rmsd) for each ensemble is high (F1, rmsd = 5.83 Å; F2, rmsd = 6.44 Å) and reflects not only the limited number of NMR restraints but also the expected lability and disorder inherent within each sequence. The best structures obtained for each sequence reveal the expected broad scatter of phi, psi torsion angles (Figure S5 in the Supporting Information) as well as the disordered backbone

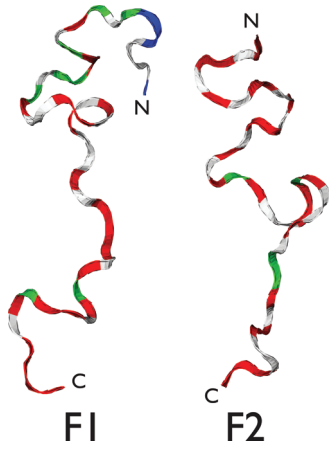

FIGURE 6: SA/MD XPLOR lowest energy backbone conformation of F1 and F2. Ribbon representation. Ribbon colors refer to residue type as follows: green = polar; red = anionic; blue = cationic; white = nonpolar.

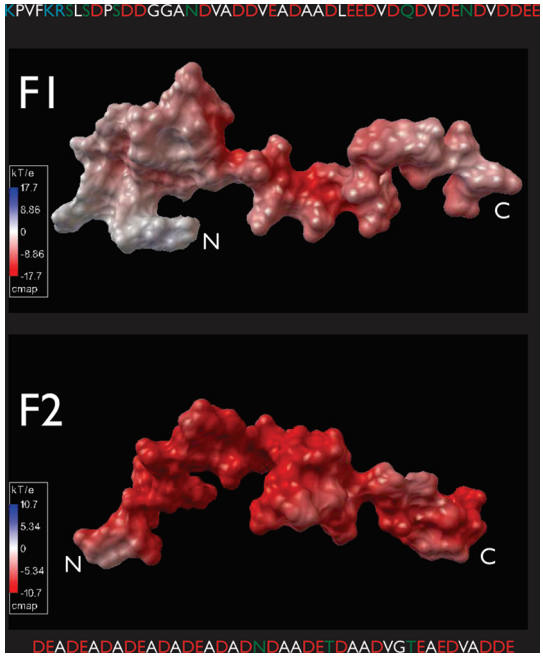

FIGURE 7: Calculated APBS solvent electrostatic molecular surface of lowest energy conformer (full atom representation) of F1 and F2. Electrostatic surface rendering is as follows: red = anionic; blue = cationic. The primary sequences for F1 and F2 are also presented, with the color coding as follows: green = polar; red = anionic; blue = cationic; white = nonpolar.

configuration induced by the combination of extended beta strand and random coil segments (Figures 5 and 6).

We calculated the Poisson−Boltzmann solvation energies (*22*) for the lowest energy conformers of F1 and F2 and generated a solvation electrostatic molecular surface for these sequences (Figure 7, full atom representation). Here, we find that the lowest energy structures have different solvation energies (F1 = −2.3287 E4 kJ/mol; F2 = −2.821 E4 kJ/mol, a difference of 18%), from which we infer F2 has a greater solvent accessibility. Further, the solvation electrostatic surfaces of both sequences differ with regard to negative charge density (Figure 7). We believe that these factors contribute to the observed differences in peptide:metal ion complex formation (Figure 2). Specifically, the anionic charge density of F2 is consistently high along most of its sequence length and is solvent-accessible. This explains why F2 exhibits higher metal ion binding capacity (Figure 2) and enhanced local conformational response to Ca(II) (Figures 3 and 4). An interesting situation exists in F1, where the KPVFKRS region generates a cationic molecular surface immediately adjacent to the partially anionic surface of the −LSDPSDDGGA− region. The close proximity of opposite side chain charge leads to neutralization and partial structural collapse, with reduced solvent accessibility and the potential for reduced metal ion complexation at the N-terminus (Figure 7). Thus, we believe that the significant region of anionic surface charge density and metal ion complexation in F1 is within the N18 to E48 region (Figure 7).

DISCUSSION

Our present work extends the current knowledge of polyelectrolyte domain structure and organization within IDPs and biomineralization proteins. The Asprich family is a disordered sequence clan that possess two common disordered regions (Figure 1). The first is a short segment located near the N-terminus of these proteins and exhibits overlap with the synthetic F1 sequence fragment (*8*). The second and major contributor to disorder is the Asp, Glu rich domain that initiates near residue 60 (Figure 1). Interestingly, a third, 5 AA disordered region was identified near the C-terminus in Asprich "a" through "g" and this small sequence overlaps with the −EDVAD− region of the synthetic F2 fragment (*8*). Notably, this C-terminal domain is absent from the sequences of Asprich "1" through "3" (Figure 1). We take this to mean that this highly anionic, partially disordered C-terminal region must play a distinguishing role in the function of Asprich "a" through "g" but is unnecessary for the function of "1" through "3".

Our research demonstrates that the conserved N- and C-terminal sequences of Asprich "a" through "g" are globally disordered and readily form complexes with metal ion "targets" (Figure 2) without incurring significant global conformational rearrangement (Figure S3 in the Supporting Information). The reported effect of metalation on IDPs is somewhat variable; on the one hand, IDPs such as retinal phosphodiesterase gamma subunit (*23a*) are structurally unaffected by metals, whereas other IDPs such as prothymosin alpha (*23c*) and alpha synuclein (*23b*) undergo folding in the presence of metal ions with functional consequences. It is likely that protein−metal ion binding affinity plays a role in these instances. We conclude that the absence of a metalation effect is functionally relevant for the disordered terminal sequences of Asprich "a" through "g". Most likely, the inability of F1 and F2 to fold in the presence of excess Ca(II) is linked to the metal ion binding affinities within these disordered sequences, and studies are currently underway to assess this.

We note that F1 and F2 exhibit different structural and sequence features that are potentially important for their function. For example, in addition to a highly anionic sequence stretch (N18−E48), F1 possesses a partially collapsed, electrostatically neutral N-terminal region (Figure 7) that experiences slow interconversion between two conformational states (Figure 3). We believe that the N-terminal region of F1 does not serve as an effective metal ion binding site. In contrast, F2 possesses a higher anionic surface density and solvent accessibility (Figure 7), and this explains why

F2 forms higher order metal ion adducts (Figure 2; Figures S1 and S2 in the Supporting Information) and exhibits a more significant local conformational response to Ca(II) ions (Figures 3 and 4). These structural and electrostatic differences may explain the different abilities of F2 and F1 to interact with calcite surfaces and step edges and modulate crystal growth kinetics and morphology modulation (*8*).

With the identification of the Asprich protein family as an IDP member, we now extend the functional scope of IDP proteins from participation in intracellular activities such as translational and transcriptional regulation (*2, 3, 32*) and into the realm of extracellular biomineral formation. We suspect that extended beta strand and coil segments, enhanced metal ion binding, and solvation effects found in Asprich sequences may also be common to other Asp, Glu polyelectrolyte domains within other IDP proteins and potentially drive the function of these domains. Although the specific functions of polyelectrolyte domains in other IDP members cannot be addressed by our present study, we can speculate as to how the N- and C-terminal sequences assist Asprich in the induction and stabilization of ACC and inhibition of calcite nucleation (*10*). Given that disorder-to-order transitions in IDP proteins are coupled with binding promiscuity (*2, 32*), the presence of disordered regions and their inherent flexibility within Asprich proteins would permit conformational adaptation of these proteins to biomineralization "targets" that offer a mixture of surface features, such as nucleation clusters (*8*), ACC particles, cell surface receptors (*2, 3*), other matrix proteins, or calcite terraces or growth step regions (*8*). Certainly, in the case of ACC clusters, the absence of long-range backbone ordering within the terminal Asprich sequences would complement the amorphous, short-range ordering and irregular topography expected on ACC surfaces (*10*). What is lacking at this point is evidence of a global disorder-to-order transformation that should accompany events such as Asprich binding to ACC, to calcite terraces or steps, or the assembly of nucleation clusters on Asprich molecular surfaces. These issues will be addressed in subsequent experiments.

SUPPORTING INFORMATION AVAILABLE

Ion trap mass spectrometry time-of-flight spectra and adducts of fragments 1 and 2 in the presence of $CdCl_2$, $LaCl_3$, $EuCl_3$ (Figures S1, S2; Tables S1, S2); circular dichroism spectrometry of fragments 1 and 2 in the presence of $CaCl_2$ (Figure S3); aligned backbone structures of fragments 1 and 2 conformer ensembles obtained from NMR restrained simulated annealing molecular dynamics (Figure S4); and Ramachandran ϕ, ψ dihedral angle distribution plots for best energy structures of fragments 1 and 2 (Figure S5). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. (a) Dyson, H. J., and Wright, P. E. (1998) Equilibrium NMR studies of unfolded and partially folded proteins. *Nat. Struct. Biol. 7*, 499–503. (b) Dyson, H. J., and Wright, P. E. (1998) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol. 339*, 258–270.
2. (a) Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci. 11*, 739–756. (b) Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci. 27*, 527–533. (c) Meszaros, B., Tompa, P., Simon, I., and Dosztanyi, Z. (2007) Molecular principles of the interactions of disordered proteins. *J. Mol. Biol. 372*, 549–561. (d) Tran, H. T., Mao, A., and Pappu, R. V. (2008) Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc. 130*, 7380–7382.
3. Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry 44*, 1989–2000.
4. Kapłon, T. M., Rymarczyk, G., Nocula-Ługowska, M., Jakób, M., Kochman, M., Lisowski, M., Szewczuk, Z., and Ozýhar, A. (2008) Starmaker exhibits properties of an intrinsically disordered protein. *Biomacromolecules 9*, 2118–2125.
5. Delak, K., Harcup, C., Lakshminarayanan, R., Sun, Z., Fan, Y., Moradian-Oldak, J., and Evans, J. S. (2009) The tooth enamel protein, amelogenin, is an intrinsically disordered protein with an extended molecular configuration in the monomeric form. *Biochemistry 48*, 2272–2281.
6. Amos, F. F., and Evans, J. S. (2009) AP7, a partially disordered pseudo C-RING protein, is capable of forming stabilized aragonite in vitro. *Biochemistry 48*, 1332–1339.
7. (a) Evans, J. S. (2008) "Tuning in" to mollusk shell nacre- and prismatic-associated protein terminal sequences. Implications for biomineralization and the construction of high performance inorganic-organic composites. *Chem. Rev. 108*, 4455–4462. (b) Wustman, B. A., Weaver, J. C., Morse, D. E., and Evans, J. S. (2003) Characterization of a Ca(II)-, mineral-interactive polyelectrolyte sequence from the adhesive elastomeric biomineralization protein Lustrin A. *Langmuir 19*, 9373–9381. (c) Wustman, B. A., Weaver, J. C., Morse, D. E., and Evans, J. S. (2002) Structural analyses of polyelectrolyte sequence domains within the adhesive elastomeric biomineralization protein Lustrin A. *Langmuir 18*, 9901–9906.
8. (a) Collino, S., Kim, I. W., and Evans, J. S. (2006) Identification of an "Acidic" C-Terminal mineral modification sequence from the mollusk shell protein Asprich. *Cryst. Growth Des. 6*, 839–842. (b) Kim, I. W., Giocondi, J. L., Orme, C., Collino, S., and Evans, J. S. (2008) Morphological and kinetic transformation of calcite crystal growth by prismatic-associated Asprich sequences. *Cryst. Growth Des. 8*, 1154–1160. (c) Delak, K., Giocondi, J., Orme, C., and Evans, J. S. (2008) Modulation of crystal growth by the terminal sequences of the prismatic-associated Asprich protein. *Cryst. Growth Des. 8*, 4481–4486.
9. Gotliv, B.-A., Kessler, N., Sumerel, J. L., Morse, D. E., Tuross, N., Addadi, L., and Weiner, S. (2005) Asprich: A novel aspartic acid rich protein family from the prismatic shell matrix of the bivalve Atrina rigida. *ChemBioChem 6*, 304–314.
10. Politi, Y., Mahamid, J., Goldberg, H., Weiner, S., and Addadi, L. (2007) Asprich mollusk shell protein: In vitro experiments aimed at elucidating function in $CaCO_3$ crystallization. *CrystEngComm 9*, 1171–1177.
11. Sethi, D., Garg, A., and Raghava, G. P. S. (2008) DPROT: Prediction of disordered proteins using evolutionary information. *Amino Acids 35*, 599–605.
12. Linding, R., Russell, R. B., Neduva, V., and Gibson, T. J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res. 31*, 3701–3708.
13. (a) Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Informatics 10*, 30–40. (b) Romero, P., Obradovic, Z., Li, X., Garner, E., Brown, C., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins: Struct., Funct., Genet. 42*, 38–48. (c) Romero, P., Obradovic, Z., and Dunker, A. K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Informatics 8*, 110–124.
14. Yi, S., Boys, B. L., Brickenden, A., Konermann, L., and Choy, W. Y. (2007) Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: Evidence for metalation as an entropic switch. *Biochemistry 46*, 13120–13130.
15. Li, M., Liu, J., Ran, X., Fang, M., Shi, J., Qin, H., Goh, J. M., and Song, J. (2006) Resurrecting abandoned proteins with pure water: CD and NMR studies of protein fragments solubilized in salt-free water. *Biophys. J. 91*, 4201–4209.
16. (a) Collino, S., and Evans, J. S. (2008) The molecular specifications of a mineral modulation sequence derived from the aragnote-promoting protein, n16. *Biomacromolecules 9*, 1909–1918. (b) Collino, S., Kim, I. W., and Evans, J. S. (2008) Identification and

structural characterization of an unusual RING-like sequence within an extracellular biomineralization protein, AP7. *Biochemistry 47*, 3745–3755.
17. Collino, S., and Evans, J. S. (2007) Structural features that distinguish kinetically distinct biomineralization polypeptides. *Biomacromolecules 8*, 1686–1694.
18. Wang, Y., Nip, A. M., and Wishart, D. S. (1997) A simple method to quantitatively measure polypeptide $J_{HNH\alpha}$ coupling constants from TOCSY or NOESY spectra. *J. Biomol. NMR 10*, 373–382.
19. MacKerell, A. D., Brooks, B., Brooks, C. L., Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998) CHARMM: The Energy Function and Its Parameterization with an Overview of the Program, in *The Encyclopedia of Computational Chemistry* (Schleyer, P. v. R., et al., Eds.) Vol. 1, pp 271−277, John Wiley & Sons, Chichester.
20. (a) Kumar, A., Wagner, G., Ernst, R. R., and Wuthrich, K. (1981) Buildup rates of the nuclear Overhauser effect measured by two-dimensional proton magnetic resonance spectroscopy: Implications for studies of protein conformations. *J. Am. Chem. Soc. 103*, 3654–3658. (b) Wuthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley-Interscience, New York, NY. (c) Braun, W., and Go, N. (1985) Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J. Mol. Biol. 186*, 611–626.
21. Wüthrich, K., Billeter, M., and Braun, W. (1993) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol. 169*, 949–961.
22. Baker, N. H., Sept, D., Joseph, S., Holst, N. J., and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci U.S.A. 98*, 10037–10041.
23. (a) Uversky, V. N., Permyakov, S. E., Zagranichny, V. E., Rodionov, I. L., Fink, A. L., Cherskaya, A. M., and Permyakov, E. A. (2002) Effect of zinc and temperature on the conformation of the alpha subunit of retinal phosphodiesterases: A natively unfolded protein. *J. Proteome Res. 1*, 149–159. (b) Uversky, V. N., Li, J., and Fink, A. L. (2001) Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. *J. Biol. Chem. 276*, 44284–44296. (c) Uversky, V. N., Gillespie, J. R., Millett, I. S., Khodyakova, A. V., Vasilenko, R. N., Vasiliev, A. M., Rodionov, I. L., Kozlovskaya, G. D., Dolgikh, D. A., Fink, A. L., Doniach, S., Permyakov, G. D., and Abramov, V. M. (2000) Zn(II)-mediated structure formation and compaction of the "natively unfolded" human prothymosin alpha. *Biochem. Biophys. Res. Commun. 267*, 663–668. (d) Munishkina, L. A., Fink, A. L., and Uversky, V. N. (2004) Conformational prerequisites for formation of amyloid fibrils from histones. *J. Mol. Biol. 342*, 1305–1324.
24. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Biomol. NMR 12*, 1–23.
25. Ma, K., San, L. S., and Wang, K. (2001) Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. *Biochemistry 40*, 3427–3438.
26. (a) Serrano, L. (1995) Comparison between the phi distribution of the amino acids in the protein database and NMR data indicates that amino acids have various phi propensities in the random coil conformation. *J. Mol. Biol. 254*, 322–333. (b) Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996) Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol. 255*, 494–506.
27. Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J., and Dobson, C. M. (1996) Towards a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem. 100*, 2661–2666.
28. Eker, F., Griebenow, K., Cao, X., Nafie, L. A., and Schweitzer-Stenner, R. (2004) Tripeptides with ionizable side chains adopt a perturbed Polyproline II structure in water. *Biochemistry 43*, 613–621.
29. (a) Sarkar, P., Reichman, C., Saleh, T., Birge, R., and Kalodimos, C. G. (2007) Proline cis-trans isomerization controls autoinhibition of a signalling protein. *Mol. Cell 25*, 413–426. (b) Andreotti, A. H. (2003) Native state proline isomerization: An intrinsic molecular switch. *Biochemistry 42*, 9515–9524. (c) Nicholson, L. K., and Lu, K. P. (2007) Prolyl cis-trans isomerization as a molecular timer in Crk signaling. *Mol. Cell 25*, 483–486.
30. Kulp, J. L., Shiba, K., and Evans, J. S. (2005) Probing the conformational features of a phage display polypeptide sequence directed against single-walled carbon nanohorn surfaces. *Langmuir 21*, 11907–11914.
31. Pai, R. K., and Pillai, S. (2008) Divalent cation-induced variations in polyelectrolyte conformation and controlling calcite morphologies: Direct observation of the phase transition by atomic force microscopy. *J. Am. Chem. Soc. 130*, 13074–13078.
32. Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Walton, J., Vacic, V., Obradovic, Z., and Uversky, V. N. (2008) The unfoldomics decade: An update on intrinsically disordered proteins. *BMC Genomics 9*, 1–26.

BI900113V