The Tooth Enamel Protein, Porcine Amelogenin, Is an Intrinsically Disordered Protein with an Extended Molecular Configuration in the Monomeric Form[†]

Katya Delak,[‡] Craig Harcup,[‡] Rajamani Lakshminarayanan,[§] Zhi Sun,[§] Yuwwei Fan,[§] Janet Moradian-Oldak,*[,§] and John Spencer Evans*[,‡]

Laboratory for Chemical Physics, New York University, 345 East 24th Street, Room 1007, New York, New York 10010, and Center for Craniofacial Biology, CSA 107, School of Dentistry, Health Sciences Center, University of Southern California, Los Angeles, California 90033

Received November 25, 2008; Revised Manuscript Received January 12, 2009

ABSTRACT: Amelogenins make up a class of proteins associated with the formation of mineralized enamel in vertebrates, possess highly conserved N- and C-terminal sequence regions, and represent an interesting model protein system for understanding biomineralization and protein assembly. Using bioinformatics, we report here the identification of molecular traits that classify 12 amelogenin proteins as members of the intrinsically disordered or unstructured protein family (IDPs), a group of proteins that normally exist as unfolded species but are capable of transformation to a folded state as part of their overall function. Using biophysical techniques (CD and NMR), we follow up on our bioinformatics studies and confirm that one of the amelogenins, recombinant porcine rP172, exists in an extended, unfolded state in the monomeric form. This protein exhibits evidence of conformational exchange between two states, and this exchange may be mediated by Pro residues in the sequence. Although the protein is globally unfolded, we detect the presence of local residual secondary structure [α-helix, extended β-strand, turn/loop, and polyproline type II (PPII)] that may serve several functional roles within the enamel matrix. The extended, labile conformation of rP172 amelogenin is compatible with the known functions of amelogenin in enamel biomineralization, i.e., self-assembly, associations with other enamel matrix proteins and with calcium phosphate biominerals, and interaction with cell receptors. It is likely that the labile structure of this protein facilitates interactions of amelogenin with other macromolecules or with minerals for achievement of internal protein stabilization.

The formation of inorganic compounds by organisms (biomineralization) is a substantial scientific puzzle. The ability of cells to employ proteins to control nucleation, crystal morphology, polymorphism, and the material properties of living tissues requires precise molecular control and efficient mechanisms (*1*). One such protein, amelogenin, is found in mammalian tooth enamel, one of the most highly mineralized materials of vertebrates (*1*−*3*). Amelogenin is essential for normal enamel development and is capable of protein self-association, forming supramolecular assemblies under defined conditions in the laboratory (*4*−*6*). These supramolecular assemblies (nanospheres) are believed to exert control over the morphology, organization, and directionality of hydroxyapatite crystal growth (*7, 8*). Primary sequence analysis of 26 mammalian lineages indicates that the N-terminus (Tyr-rich) and C-terminus (charged) of amelogenin are highly conserved, whereas variations occur in the central regions (*9*). Amelogenin sequence mutations lead to defective enamel crystal formation and organization (*10, 11*), and deletion of the conserved terminal domains leads to the formation of ill-defined enamel crystals, highlighting the importance of these conserved domains in protein−protein or protein−mineral interactions (*12*). It has been suggested that amelogenin can interact with other important enamel matrix proteins that participate in enamel formation (*13*). Hence, the amelogenin protein family represents a highly relevant model for understanding protein-mediated biomineralization processes in nature.

One of the keys to understanding amelogenin function is to establish the structure of this protein. However, the determination of amelogenin structure has been limited by both the intrinsic lability of this protein and the self-association phenomenon. These traits are also common to other unfolded proteins, and recent studies have established that unfolded domains are functionally important for these proteins (*14*−*19*). Proteins or domains that lack organization and exist as a dynamic ensemble of interconverting structures have been classified as intrinsically disordered or natively unfolded proteins (IDPs)[1] (*18, 19*). The amino acid compositions of IDPs are less complex and are dominated by

[†] This work was supported by funding from the Department of Energy (DE-FG02-03ER46099) and the National Institute of Dental and Craniofacial Research (DE-013414).

* To whom correspondence should be addressed. J.S.E.: e-mail, jse1@nyu.edu; telephone, (212) 998-9605; fax, (212) 995-4087. J.M.-O.: e-mail, joldak@usc.edu; telephone, (323) 442-1759; fax, (323) 442-2981.

[‡] New York University.

[§] University of Southern California.

[1] Abbreviations: UDDW, unbuffered deionized distilled water; IDP, intrinsically disordered or unfolded protein; CD, circular dichroism; rP172, recombinant porcine amelogenin (172 amino acids); PPII, polyproline type II; rmsd, root-mean-square deviation.

10.1021/bi802175a CCC: $40.75
Published on Web 02/23/2009

disorder-promoting residues (E, K, R, G, Q, S, P, and A) (*19*). The functional attributes of IDPs include self-assembly into organized structures or binding to multiple targets with concomitant structural alterations (*18, 19*). These compositional and functional attributes are also common to the amelogenins as well (*3*−*9, 20*−*23*).

Hence, we believe that the functions of amelogenin will be better understood once we understand the molecular behavior of the unfolded, preassociative monomeric form of this protein. Such information will allow us to clarify conformational transitions and nonbonding interactions that occur when amelogenin interacts with itself or with other enamel matrix targets (i.e., non-amelogenins, cell surfaces, and mineral phase). This can be achieved by employing methods such as bioinformatics (*24, 25*) and biophysical techniques (*14*−*17, 21*) to directly link amelogenin to the IDP protein family and establish folded and unfolded regions of this protein. For biophysical experiments to be successful, care must be taken to stabilize the monomeric form of amelogenin and avoid problems that are encountered by aggregation (*20*−*23*). Recently, it has been shown that aggregation problems in proteins can be successfully minimized by utilizing unbuffered deionized distilled water (UDDW), where buffer-induced salting out effects that drive hydrophobic interactions are negligible (*26*). Under these conditions, a high degree of protein solubility, minimal hydrophobic intermolecular interactions, excellent NMR signal dispersion, and the appearance of important intrinsic conformational states have been reported (*26*). Hence, with this approach, a joint investigation using bioinformatics and IDP-specific structural techniques such as NMR spectroscopy (*14*−*18*) is now more feasible.

This study establishes a link between amelogenin and the IDP protein class. We first employ bioinformatic approaches (*24, 25*) to fingerprinting amelogenins as IDPs. We then utilize NMR spectroscopy, an IDP "friendly" technique (*14*−*18, 26*), to determine the global and local conformational states of a specific amelogenin (recombinant porcine amelogenin, rP172) in solution. Our bioinformatics data confirm that the majority of amelogenins from different species fall within the IDP category. This conclusion is also supported by our NMR data, where we observe that monomeric porcine rP172 amelogenin exists in an extended molecular configuration that consists of random coil conformation and partially structured regions (PPII, helix, extended β-strand, turn, or loop). These molecular features are consistent with both the in vitro behavior and the known in vivo enamel matrix functions of amelogenins.

MATERIALS AND METHODS

Calculation of Mean Hydrophobicity and Net Charge. The normalized values of the Kyte and Doolittle hydrophobicity scale for individual residues were obtained from http://www.expasy.org/tools/protscale.html. The amino acid sequences of the intrinsic disorder and folded proteins were obtained from ref *24*. The mean hydrophobicity is the sum of the hydrophobicity of all amino acid residues divided by the total number of residues, and the mean net charge is the absolute net charge at pH 7.0 divided by the total number of residues.

Preparation of rP172 Recombinant Amelogenin. Recombinant porcine amelogenin [U-^{13}C,^{15}N]rP172 was expressed in Bio-Express Cell Growth Media (U-^{13}C, 98%, U-^{15}N, 98%) using *Escherichia coli* strain BL21-codon plus (DE3-RP, Stratagene, La Jolla, CA) as previously described (*3, 4*). The protein in the cell lysate was precipitated by 20% ammonium sulfate and purified using a reversed-phase column, in a Varian Prostar HPLC system. The precipitate was dissolved in 0.1% TFA and loaded onto a C4 semi-preparative column (214TP510, Vydac, Hesperia, CA) and fractionated using a linear gradient of 60% (v/v) acetonitrile at a flow rate of 2 mL/min. The homogeneity of the protein was confirmed by analytical chromatography using a C4 analytical column (214TP54, Vydac). The purity of the protein was more than 98%. rP172 has 172 amino acids and is an analogue of native P173 porcine amelogenin without the first methionine and the phosphate on Ser16 (*27*).

Dynamic Light Scattering Experiments. DLS measurements were performed on a DynaPro-99EMS/X instrument equipped with a solid-state laser operating at 655 nm with a temperature-controlled MicroSampler at 10 °C (Wyatt Technologies, Santa Barbara, CA). Amelogenin rP172 was dissolved in unbuffered double-distilled water (UDDW) to a final concentration of 1.5 mg/mL (75 μM, pH 6.0). The normalized intensity correlation function was analyzed by a regularization method included in the data analysis software package (Dynamics version 6.3.01), to give the information about the distribution of the exponential decay function with decay rate, Γ. The acquisition time for each run was set at 10 s. The experiment continued for 50−60 runs which corresponded to 8−10 min. The data analysis software analyzed the correlation function and provided the hydrodynamic radii and mass distribution of particles.

Circular Dichroism Experiments. CD measurements for amelogenin in UDDW were conducted on a JASCO J-810 spectropolarimeter calibrated using a 0.06% (+)-10-camphorsulfonic acid solution. The instrument optics and sample chamber were continually flushed with dry N_2 gas at a rate of 20 L/min. The CD spectra were recorded at 10 °C in a 1 mm path length quartz cell (300 μL) by using a scanning speed of 50 nm/min, a response time of 1 s, a bandwidth of 1 nm, and an average of eight scans.

The variable-temperature (VT) experiments were performed on this same spectropolarimeter using a Peltier setup. For these VT experiments, the instrument optics and sample chamber were continually flushed with dry N_2 gas at a rate of 20 L/min. The CD spectra were recorded at 5 and 10 °C in a 1 mm path length quartz cell (300 μL) by using a scanning speed of 50 nm/min, a response time of 1 s, a bandwidth of 1 nm, and an average of eight scans. A total of 16 scans were recorded to reduce the signal-to-noise ratio, and the resultant spectra were baseline-subtracted, and smoothed by the Savitzky−Golay method. The thermal dependency of the ellipticity at 224 nm was monitored at a heating or cooling rate of 60 °C/h.

NMR Spectroscopy. All experiments were performed at 283 K on a Bruker AVANCE 700 MHz four-channel NMR system equipped with an x,y,*z* axis PFG 5 mm HXY cryoprobe. The low temperature was chosen to stabilize the sample against aggregation and to slow conformational exchange. A rP172 sample was created using lyophilized [U-^{13}C,^{15}N]rP172 dissolved in Milli-Q pure deionized distilled water and 10% (v/v) D_2O (99.99 atom % D, Cambridge Isotope Laboratories) in the absence of any buffer (i.e., the

protein acts as its own buffer). This creates a 75 μM, pH 3.8 sample as measured with a NMR pH microelectrode using a standardized Radiometer Copenhagen pH meter. Monitoring of sample pH was performed periodically between experiments, and it was found that the sample pH remained stable near 4. Samples were placed in 5 mm symmetrical D_2O-matched Shigemi NMR microtubes (Shigemi, Inc., Alison Park, PA). No visible aggregation was evident, and periodic ^{15}N HSQC experiments were conducted to verify that line width broadening or chemical shift changes associated with aggregation did not occur. Sequential assignments were obtained using a suite of multidimensional NMR experiments (backbone-specific, HNCO, HNCACO, CBCACONH, and HNCACB; side chain-specific, CCCONH and HCCCONH) (*15, 16*). In addition to the backbone and side chain experiments, $^{15}N-^{1}H$ HSQC NOESY experiments (*16*) were conducted at optimal mixing times of 25, 50, 75, and 100 ms. NOE assignments were made using Sparky visualization software (SPARKY 3, University of California, San Francisco, CA). HNHA experiments (*28*) were conducted to determine $^{3}J_{CH\text{-}NH}$ coupling constants, and corresponding dihedral ϕ angles were calculated using the equation in the accompanying reference. All data were processed using NMRPipe (National Institutes of Health, Bethesda, MD). The spectra were referenced with respect to the temperature-corrected water resonance (*29*), and ^{13}C and ^{15}N chemical shifts were referenced on the basis of the ^{1}H reference following IUPAC guidelines using the unified chemical shift scale (*29, 30*). Acquisition and processing parameters are provided in the appropriate figure legends and Supporting Information. Sequential assignments were made using NMRViewJ, version 8.0.a11 (One Moon Scientific, Inc., Newark, NJ).

rP172 Amelogenin Model Building. The initial rP172 amelogenin Protein Data Bank files were generated using MacPyMol (DeLano Scientific, Palo Alto, CA). Torsion angles of the initial structures were not deliberately specified. Two structures were generated: one in which a neutral pH scenario was envisioned, where aspartate and glutamate side chains were negatively charged, and another assuming an acidic environment, in which the aspartate and glutamate side chains were protonated. In both cases, all cationic (Arg, His, and Lys) side chain groups were positively charged and the terminal regions were represented as NH_3^+ and COO^-. All calculations were carried out using XPLOR-NIH. Protein parameters were defined using the default XPLOR protein parameter file, protein.par, which is based on the CHARMM19 force field. NMR interproton distance restraints based on experimental NOESY data were generated using the noe2xplor.py Python script (http://pbil.ibcp.fr/~nsapay/tools/script.html). Peak intensities were categorized as strong (1.8−2.8 Å), medium (1.8−3.6 Å), weak (1.8−5.0 Å), and very weak (1.8−6.0 Å). Pseudoatom corrections were used where necessary, and the corresponding corrections were included as part of the restraint input file. 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. A total of 239 NOE restraints were used for the calculations. In addition to NOE distance restraints, residue-specific backbone dihedral angle restraints (ϕ and ψ) were used as part of the simulations. These restraints were generated from $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$, ^{15}N, and ^{13}CO using TALOS. A total of 149 dihedral angle restraints were generated. The molecules were solvated implicitly through the use of a distance-dependent dielectric constant of 78.5.

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 ϕ and ψ angle. This was followed by 30 ps of equilibrium dynamics at 300 K that were added after the initial minimization and before the structure was heated to 2000 K. The system was cooled from 2000 K at 100 K per step to 300 K. The resultant structures were then energy minimized with the Powell method. Calculations were conducted 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 coefficient. The accept.inp script was used to search the 1000 calculated structures for the 100 lowest-energy structures with zero NOE restraint violations and a minimum number of dihedral angle violations. From these 100 structures, the 10 lowest-energy structures were chosen for a conformer library; the final lowest-energy structure was then identified for additional comparisons. Visualization of structures, Ramachandran maps, and backbone alignment and rmsd calculations were performed using the VMD 1.8.6 software package.

Table 1: Analysis of Disorder- and Order-Promoting Residues in Amelogenin Sequences

amelogenin (species)	% disorder-promoting residues	% order-promoting residues
bovine	56.9	25.4
chimpanzee	56	24.6
guinea pig	55.2	27.1
goat	53.8	26.7
golden hamster	58.3	24.4
horse	52.3	31.3
human	56	24.6
macaque	55.4	24.6
marmoset	55.7	24.4
orangutan	55.4	24.6
pig	54.9	25.4
rat	58.3	24.4

RESULTS

Amino Acid Sequence Analysis of Vertebrate Amelogenins. Mammalian amelogenin sequences are highly conserved, and for the purposes of this study, we chose to analyze the sequences from 12 representative species. Interestingly, all 12 vertebrate amelogenins contain 50−60% disorder-promoting residues (Table 1) (*19*). In the absence of experimental assignments, the mean net charge ($\langle R \rangle$) versus mean hydrophobicity ($\langle H \rangle$) correlation (CH plot) has been developed for binary classification of proteins and provides perhaps the most intuitive description of overall protein conformation (*18, 23*). As shown in the CH plot (Figure 1), the IDPs are separated from folded proteins (indicated by the line) by the relation $\langle H \rangle_{boundary} = (\langle R \rangle + 1.151)/2.785$. In the CH plot, porcine amelogenin is located in the region of intrinsically disordered proteins.

Biophysical Characterization of Amelogenin rP172 in Pure Water. Confirming the observations obtained for other aggregation-prone proteins (*26*), we found recombinant porcine amelogenin rP172 to be highly soluble in unbuffered,

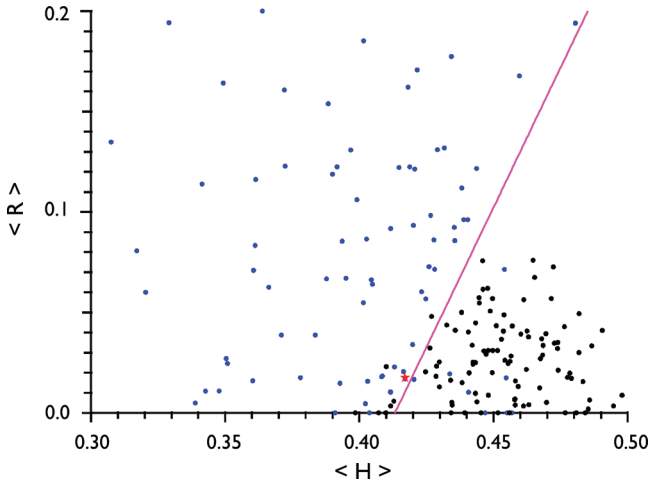

FIGURE 1: Computational analysis (CH plot) of porcine amelogenins. The mean net charge ($\langle R \rangle$) vs mean hydrophobicity ($\langle H \rangle$) plot for IDPs (blue circles), folded proteins (black circles), and the boundary [magenta line; $\langle H \rangle_{boundary} = (\langle R \rangle + 1.151)/2.785$] that demarcates the two classes of proteins. The red star indicates the position of porcine amelogenin in the CH plot.

Table 2: Dynamic Light Scattering Data for rP172 Amelogenin in UDDW[a]

item	R (nm)	polydispersity	MW-R (kDa)	scattering intensity	percentage mass
peak 1	1.4	0	7	1.5	100
peak 2	103.2	15	173423	76.1	0
peak 3	587.5	13.7	10140000	13.2	0
peak 4	7970.4	0	4.53×10^9	9.2	0

[a] Regularization data representing particle size distribution of unlabeled amelogenin rP172 in water at a concentration of 1.5 mg/mL (75 μM) and 10 °C. Note that unlabeled amelogenin rP172 was used. R is the particle radius; MW-R is the molecular weight. While the percentage of scattering intensity is mostly contributed by the larger aggregates, the majority of the mass is contributed by the monomers of amelogenin (percentage mass = 100). Amelogenin aggregates are present in trace amounts, but their scattering is very large.

deionized distilled water (UDDW) at 10 °C; the protein solution appeared transparent and exhibited no observable turbidity. When the sample was dissolved in UDDW, the final pH of the unbuffered amelogenin sample was weakly acidic (i.e., the pH ranges from 3.8 to 6.0 but is stable over time). We believe that the variations in pH from sample to sample can be attributed to a number of factors, such as the source of deionized distilled water and variations in residual buffer salt retained by each protein preparation after purification (*26*).

Dynamic light scattering analysis of an amelogenin solution (concentration of 75 μM in UDDW, pH 6.0, 10 °C) revealed the monomeric state of the protein, although trace amounts of aggregates were also detected (Table 2). The CD spectra of rP172 were recorded at 5 and 10 °C (Figure 2A), and both spectra exhibit an intense negative minimum around 202 nm with an isoelliptic point around 209 nm. For the sake of comparison, we note that IDPs exhibit characteristic CD spectra with an intense negative minimum around 195−203 nm (*31*). Strong absorption bands in the n−π* region are typically associated with the presence of an ordered secondary structure (*21, 31, 32*), and the absence of these bands for rP172 in UDDW suggests that amelogenin exists in a disordered conformation (*21, 25, 26, 31, 32*).

The presence of an isoelliptic point in the CD spectra for

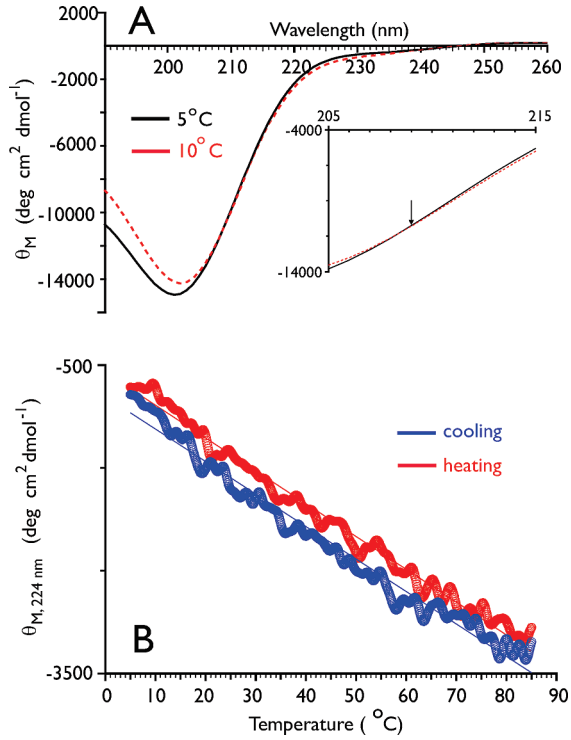

FIGURE 2: (A) CD spectra of rP172 (0.4 mg/mL) in UDDW. The inset spectrum is an enlarged region close to the isoelliptic point (denoted by the arrow). (B) Thermal dependency of ellipticity values at 224 nm for rP172 (0.4 mg/mL) in UDDW. The red circles indicate heating cycles, while the blue circles indicate cooling cycles.

rP172 indicates that an equilibrium exists between unordered and PPII conformations (*33*). A characteristic feature of proteins or polypeptides that exhibit an unordered PPII equilibrium is the presence of a linear dependency of the ellipticity values at 222 nm on temperature, in contrast to the sigmoidal dependence exhibited by proteins with α-helix/β-strand structures (*33*). To probe this, we measured the temperature-dependent melting and annealing behavior of rP172 and monitored the ellipticity intensity at 224 nm (Figure 2B). The presence of aggregation was ruled out by the absence of sample turbidity during heating as well as the reproducibility of ellipticity curves between a 10 °C sample and a parallel sample that prior to CD measurement was first heated to 85 °C and then cooled to 10 °C (Figure S1 of the Supporting Information). In UDDW, rP172 exhibits a linear transition over 5−85 °C with a slope value of −31.6 deg cm^2 dmol^{-1} C^{-1} for heating and −31.9 deg cm^2 dmol^{-1} C^{-1} for annealing (Figure 2B). The absence of any cooperative transition within this temperature range indicates that PPII structure is present at low temperatures and transforms to unordered/β-strand structure at elevated temperatures. An estimation of the PPII content was obtained from the limiting ellipticity values based on the assumption that the linear transition is an approximation to a broad sigmoidal curve (*32, 33*). From these values, the PPII content for rP172, in pure water at 10 °C, is estimated to be 22.7 ± 1.2%.

Global Conformational Features of Monomeric rP172. Intrinsically disordered or unfolded proteins present a challenge for structure determination by NMR due to the absence of stable, unique structures and the presence of conformational averaging, resulting in compromises in chemical shift dispersion, resonance broadening, and NOE cancellation

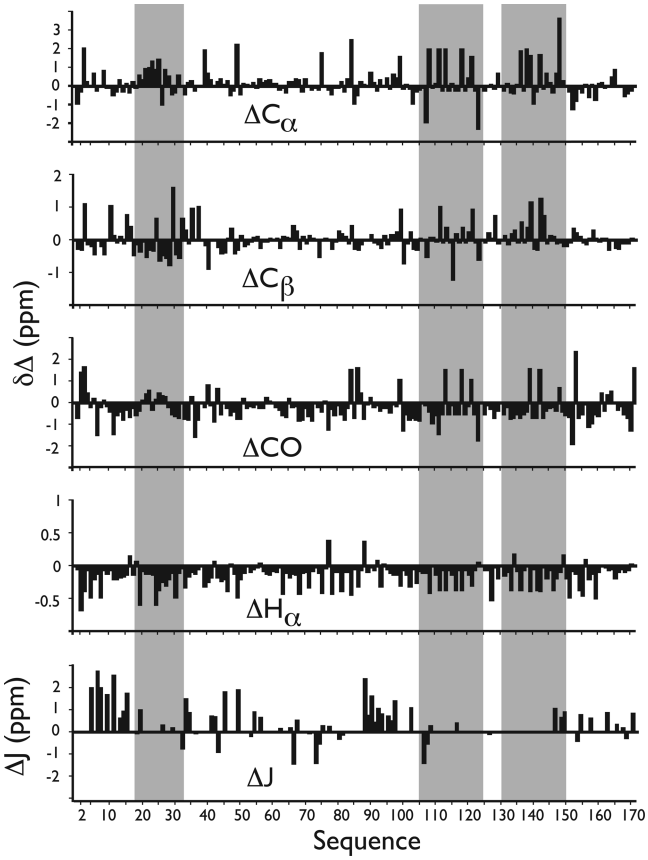

FIGURE 3: ^{13}C and ^{1}H backbone conformational shifts and ΔJ values for rP172 amelogenin in UDDW. Values were determined from the NMR data presented in Table S1 of the Supporting Information. Gray stripes indicate sequence regions that exhibit non-random coil characteristics. In general, (+) secondary shifts for $^{13}C\alpha$ and ^{13}CO indicate helical structures whereas (−) shifts indicate a propensity for extended conformations such as β-strand (*36, 37*). For $^{1}H\alpha$ and $^{13}C\beta$, (−) values indicate helical conformations and (+) values indicating extended conformations (*36, 37*). Deviations of ≥1 Hz from database $^{3}J_{CH\text{-}NH}$ random coil values are characteristic of structured domains (*37, 38*).

(*14*−*17, 26*). Following the protocol utilized in the study of other unfolded proteins (*14*−*17*), we utilized a suite of triple-resonance NMR experiments (see Materials and Methods) that take advantage of the comparatively wide chemical shift dispersions of ^{15}N and ^{13}C spins and analyzed the backbone structure of rP172. We utilized low temperatures (10 °C) to slow conformational exchange (*26, 34, 35*) and prevent aggregation. Using this approach, the majority of the backbone resonances in [U-^{13}C,^{15}N]rP172 were resolved and sequentially assigned except for E40, Q126, and H132, which could not be identified (Table S1 and Figures S2 and S3 of the Supporting Information).

Having obtained backbone structure-sensitive data [i.e., $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO, and $^{1}H\alpha$ backbone chemical shifts (*36, 37*); $^{3}J_{CH\text{-}NH}$ coupling constants (*38, 39*)], we next determined the extent of deviation of these rP172-specific NMR parameters from protein database random coil values (Figure 3), such that a global picture of rP172 folding could be obtained. Here, we note that the majority of secondary shift amplitudes are not particularly intense (Figure 3), indicating that there are no well-defined or stable helical or folded β-sheet secondary structures within rP172 in UDDW. We also analyzed the Pro $^{13}C\beta$ and $^{13}C\gamma$ chemical shifts (Table S1 of the Supporting Information) and found that all of these are within the canonical chemical shift ranges of both the trans and cis configuration (*40*−*42*). However, if we apply the more stringent chemical shift difference ($\delta[^{13}C\beta] - \delta[^{13}C\gamma]$) as a reference-independent descriptor of the Xaa−Pro peptide bond conformation (*40*), then we find that the majority of Pro residues adopt the trans configuration, with the exception of P41, P52, P96, P130, and P154, which resonate within the ^{13}C chemical shift range for the cis configuration (*40*). Thus, we conclude the following. (a) The global structure of rP172 in the monomeric form is largely unfolded, and (b) the majority of Pro residues in rP172 appear to adopt the trans configuration, similar to what is observed within the Pro-rich titin PEVK entropic chain segment (*37*), a member of the IDP protein class (*19*).

Evidence of Conformational Exchange within rP172 Amelogenin. Given that rP172 is largely unfolded, one would suspect that this protein would exhibit labile conformational qualities. This is confirmed by an examination of the HSQC spectra for rP172, where we note the presence of duplicated HSQC cross-peaks for residues V54, A63, W161, and T164 (Figures S2 and S3 of the Supporting Information). Duplication suggests the presence of two rP172 conformations, presumably in slow exchange with one another (*43*−*45*). Of these residues, only T164 exhibited well-defined cross-peaks whose volume integrations corresponding to the two conformations of T164 are approximately 2:1. Coincidentally, V54, A63, and W161 all lie within the proximity of an isolated proline residue (P59, P66, and P162, respectively). As reported elsewhere, duplication of HSQC resonances can arise from conformational exchange directed by cis−trans proline isomerization within proteins (*43*−*45*). Alternatively, it is known that Pro residues can function as a molecular "hinge" within protein sequence segments and influence the dynamics, conformation, and/or motion of sequence segments (*43*−*45*), which in turn could give rise to duplicated NMR cross-peaks (*46*). Thus, the lability of the rP172 protein is confirmed by the presence of conformational exchange, and it is possible that three of the conformational exchange sites may be modulated by Pro cis−trans interconversion and/or Pro-mediated sequence segment dynamics.

Assignment of Residual Secondary Structure within the rP172 Sequence. It is known that unfolded proteins can possess local regions where residual secondary structures reside (*14*−*17*). An examination of our qualitative NMR data set reveals that, although there is no well-defined, continuous regions of ordered structure, there are three short sequence segments that possess non-random coil J coupling and conformational shift deviations (Figure 3). The first is the V19−P33 segment within the tyrosine-rich amelogenin polypeptide domain (TRAP, P2−W45), whose negative $^{13}C\alpha$, $^{13}C\beta$, and $^{1}H\alpha$ conformational shifts show consistent deviations from random coil values that correlate with helical propensities. We also note significant deviations from ΔJ values for residues which flank this segment (Figure 3) (*38, 39*). The second region is the P105−L125 segment which contains the redundant Pro- and Gln-rich sequence, and the third is the M131−S151 segment which contains the redundant Pro, Met sequence. Both of these regions possess consistent (+) and (−) $^{13}C\alpha$, $^{13}C\beta$, and ^{13}CO conformational shifts and primarily (−) $^{1}H\alpha$ shifts, indicating mixed propensities for either helix or sheet structure in both sequence blocks (*36, 37*).

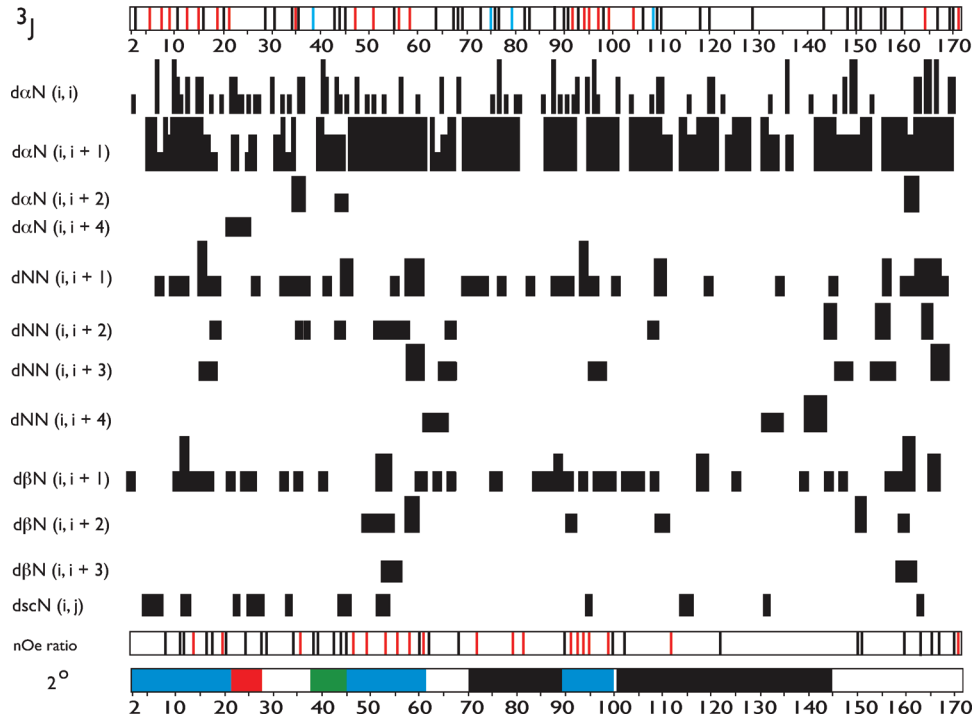

FIGURE 4: Summary of sequence-specific NMR parameters for rP172 amelogenin at pH 3.8 and 283 K in UDDW. The summary includes intraresidue NOEs; interresidue sequential, medium-range, and long-range backbone NOEs; and $^3J_{HNH\alpha}$ coupling constants (^3J). ^3J coupling constants were determined using HNHA experiments and are denoted as color-coded bars along the sequence number line [blue for α-helical (<6 Hz), red for extended chain (>8 Hz), and black for random coil (6.0−8.0 Hz)]. The absence of a bar denotes that the ^3J could not be obtained for this residue. For the sequential proton−proton NOE connectivities, the three different intensities of connecting bars indicate strong, medium, and weak NOE intensities. "nOe ratio" refers to backbone NOE intensity ratios [αN(i,i+1)/αN(i,i)] and/or [αN(i,i+1)/NN(i,i+1)] that were calculated for observable, resolvable NOE cross-peaks in the ^{15}N−^{1}H HSQC-NOESY spectra of rP172 and are denoted as color-coded bars along the sequence number line {black for random coil [αN(i,i+1)/NN(i,i+1) ≤ 1.4, and αN(i,i+1)/αN(i,i) ≤ 2.3] and red for values larger than these limits} (*41, 47*). The absence of a bar indicates that an NOE ratio could not be obtained for this residue. 2° indicates the secondary structure assignment based upon NOE and ^3J data in this figure (black for PPII, red for α-helix, blue for extended β-strand, green for β-turn or loop, and white for random coil or unstructured). The terminology and presentation of this figure follow the standardization set forth in ref *30* with some modifications.

The coexistence of global unfolded structure with residual local structure was also confirmed by quantitative NMR data [NOEs and J couplings (Figure 4)]. The absence of global folded structure is confirmed by (1) the lack of consistent sequential backbone, medium-range, and long-range backbone and sequential, medium-range, and long-range backbone−side chain NOEs (*14*−*17, 26, 47, 48*) and (2) ^3J coupling constants which predominantly fall within the random coil range (*38*). However, this data set also confirms that rP172 does possess short, dispersed sequence regions which exhibit partially ordered secondary structure traits (Figure 4). These are summarized as follows. (a) For helix, using the medium-range $d_{\alpha\text{-N}(i,i+4)}$ NOEs as a guide to α-helix occurrence (*47, 48*), we tentatively assign a short helical region at L23−Q27, which correlates with conformational shift deviations in that region (Figure 3). (b) For extended β-strand, using the extended β-strand criteria of ^3J > 8 Hz, NOE intensity ratios, and the presence of short-range $d_{\alpha\text{-N}(i,i+1)}$ and $d_{\text{NN}(i,i+1)}$ and the absence of medium-range backbone and backbone−side chain NOEs (*47, 48*), we tentatively assign extended β-strand structure to sequence regions of P2−P22, L46−H62, and G90−Q101. This correlates with the low Pro content in these three sequence segments and the large deviation in ΔJ values (Figure 3). (c) For turn or loop regions, using the criteria of sequential and medium-range $d_{\alpha\text{-N}(i,i+2)}$ and $d_{\text{NN}(i,i+2)}$ NOEs (*48*) and ^3J < 6 Hz (*38*), we assign β-turn or loop structure to the Y37−W45 segment. Not surprisingly, this sequence segment, -YGYEPMGGW-, contains turn- or loop-forming amino acids Y, G, and M (*34, 35*).

Our CD studies suggested that rP172 possesses PPII structure. Logically, we would consider that PPII structure would most likely occur within the sequence regions of I70−P89 and P102−P145, which possess a high content of the PPII-forming amino acids, Pro and Gln (*17, 49, 50*). Relative to other regions within rP172, these Pro-rich regions exhibit few if any medium-range NOEs and a low percentage of sequential nOes (Figure 4). These features have been associated with the presence of PPII secondary structure (*17, 41*). Further evidence of the presence of non-random coil structure in these two segments is found in the NOE ratios (Figure 4), where we observe that the ratios obtained for the I70−P89 region exceed random coil values. This correlates with the consistent non-random coil conformational shifts obtained for the P102−P145 region (Figure 3). On the basis of this evidence, we tentatively assign PPII structure to these two Pro regions. Thus, from NMR data, we would estimate that at 10 °C the PPII content in rP172 amelogenin is 37%, and thus, we believe that the true PPII content in this protein lies somewhere between our CD (22.7%) and NMR (37%) estimates. The remaining regions of rP172, i.e., N28−S36, A63−H69, and M146−D173, are presumed to be random coil in structure (Figure 4). Overall, our structural findings are consistent with previous biophysical studies which identified helical, sheet, and PPII structures within amelogenin in solution (*21*).

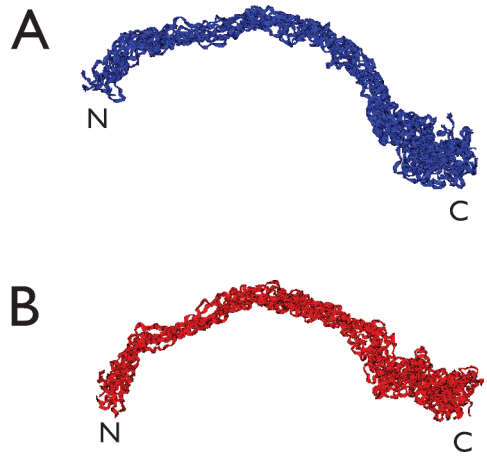

FIGURE 5: (A) Backbone-aligned ensemble (n = 10; overall backbone rmsd = 5.6 Å) of the low-energy structures obtained for rP172 amelogenin in the fully deprotonated form. (B) Same as panel A, but for the fully protonated form (n = 10; overall backbone rmsd = 5.0 Å).

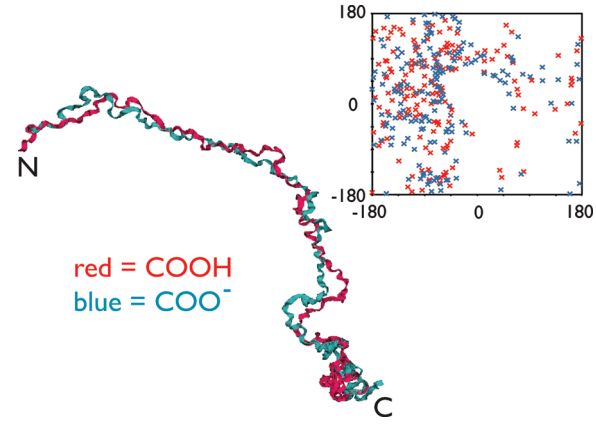

FIGURE 6: Backbone-aligned best structures obtained for the fully protonated and deprotonated forms (backbone rmsd = 4.7 Å). Ribbon representations, torsion angle plot, and rmsd backbone fitting were performed using VMD 1.8.6 for all figures. Backbone-aligned best structures obtained for the fully protonated and deprotonated forms (backbone rmsd = 4.7 Å). Ribbon representations, torsion angle plot, and rmsd backbone fitting were performed using VMD 1.8.6 for all figures. The inset is a Ramachandran plot of the XPLOR-NIH lowest-energy structures for amelogenin rP172 in the fully protonated (red) and deprotonated (blue) states.

Model Building. To gain insights into the possible molecular configurations that rP172 amelogenin adopts in the monomeric form in solution, we utilized the NOE data set along with ϕ and ψ backbone torsion angles generated by TALOS from our NMR chemical shift data set (Table S1 of the Supporting Information) as restraints for molecular dynamics simulated annealing using XPLOR-NIH. Due to the limited number of backbone restraints obtained for labile rP172 amelogenin (Figure 4), we advise that the conformations generated by our simulations should be considered qualitative in nature.

The experimental pH of the amelogenin NMR sample is 3.8, which is near the pK_a values of Asp and Glu carboxylate groups. Thus, it is likely that some percentage of the Asp and Glu residues within the protein may exist in either the deprotonated or fully protonated state. Since we cannot pinpoint which Asp and Glu residues are protonated or deprotonated within rP172, we chose to model rP172 in both the fully protonated and deprotonated states, such that the effect of Asp and Glu side chain charge representation on the global structure of the protein could be estimated. Thus, we performed parallel runs in which Asp and Glu carboxylate were represented as either fully protonated or fully deprotonated species (see Materials and Methods). A comparison of the fully protonated and deprotonated backbone aligned structural ensemble (n = 10) that fit the target NMR data set shows overall agreement in terms of global, extended conformation (Figure 5), and we note that an extended structure for amelogenin was also reported in earlier X-ray scattering studies (*51*). Thus, the representation of Asp and Glu side chain charge does not have a significant impact on the global backbone structure of rP172.

Analysis of the best or lowest-energy structures for each protonation state reveals overall good agreement in terms of backbone configuration (Figure 6). As one would expect, there are minor backbone variations which reflect differences in local chain structure in the vicinity of Asp or Glu [i.e., E18 and E40 within the tyrosine-rich amelogenin peptide (TRAP) (*52*) or A-domain (*53*)] and in the charged C-terminal region (Figure 6). This is also reflected in the Ramachandran plot for these two structures, where we note a similar distribution of backbone torsion angles within right- and left-handed helix, β-turn, and β-strand regions (Figure 6), but with slightly different values. Thus, the representation of Asp and Glu side chain charge does impact the local structure of rP172.

A comparison of the best protonated/deprotonated models of rP172 (Figure 6) with the general secondary structure features found in our NMR data set (Figure 4) shows good agreement with regard to important secondary structure features. rP172 amelogenin adopts an extended β-conformation within the P2−P22, L46−H62, and G90−Q101 regions. There is evidence of a turn or loop region near the C-terminal end of the TRAP region (Y37−W45) and extended helical or PPII-like structure in the A63−H69 and M146−D173 regions. The C-terminal domain appears to be coil-like and more condensed in configuration, which is consistent with other NMR studies of the amelogenin C-terminal region (*54*).

DISCUSSION

Our studies confirm that rP172 amelogenin is an IDP (Figures 1 and 2) that is globally extended (Figures 3 and 4) and possesses some degree of residual secondary structure, including PPII (Figures 2−6). We note that earlier studies proposed an extended configuration for amelogenin and postulated that this configuration is important for self-assembly (*51*). Our current work supports this hypothesis, and we believe that the extended, unfolded state of rP172 amelogenin offers conformational freedom (*19*) to allow interactions and multiple contacts to occur between amelogenins, with other enamel matrix proteins, or with the mineral phase (*8, 20−23, 53*). This conformational freedom is mirrored not only in the general unfolded structure of amelogenin (Figures 3−6) but also in the presence of conformational exchange (Figures S2 and S3 of the Supporting Information) where more than one protein structural state coexists in solution. Unfolded or disordered conformations also offer advantages in terms of entropy: the disorder-to-order transition that occurs upon binding of unfolded IDP to targets would decrease conformational entropy and make

highly specific interactions possible (*18, 19*). If this is true, then the interaction of rP172 amelogenin with other molecular entities would be partially driven by the need to achieve internal structural stabilization within the individual amelogenin molecules. Preliminary studies indicate that disorder-to-order transitions do occur within amelogenin during the self-assembly process in vitro (E. Beniash, personal communication), and we intend to explore this phenomenon in more detail in a subsequent report.

Although there are a diverse set of functional subcategories within the IDP class (*19*), one common trait noted for many of the IDP proteins is binding promiscuity; i.e., the protein adopts different structures upon interaction with different molecular entities (*18, 19*). We argue that the unfolded structure of rP172 amelogenin would enable a similar promiscuity in terms of structural adaptability in response to different enamel matrix targets. In turn, this promiscuity would enable rP172 amelogenin to behave in a multifunctional capacity within the enamel matrix in the following way. For example, the labile structure of amelogenin may permit interactions with cell receptor proteins that trigger amelogenin-initiated cell signaling and regulation of gene expression activities (*55, 56*). At the same time, the amelogenin protein, via its extended, unfolded N-terminal TRAP domain (amino acids 1−45), offers sufficient molecular contacts for assembly within the protein−protein interaction process (*21, 53*). Likewise, the random coil-like structure of the polar, charged C-terminal domain (amino acids 157−173) would provide enhanced multiple charged contacts with the mineral surface (*23*). Hence, the unfolded nature of monomeric amelogenin appears to be a molecular design asset, given the number of diverse activities in which this protein reportedly participates. Since assembly is the primary process associated with rP172 amelogenin, we would tentatively classify this protein as a member of the IDP "assembler" class (*19*), members of which are noted for their ability to form macromolecular complexes (*18, 19*). However, we acknowledge that additional information regarding amelogenin function may come to light at a later date, which may lead to reclassification of this protein within the IDP functional series.

Intriguing features of rP172 amelogenin are the presence and location of residual secondary structure within the protein sequence and the potential role that residual structure may play in terms of functionality. We believe that there are at least three important functions that residual structure contributes to rP172 amelogenin. First, as noted for other unfolded proteins (*14*−*17*), residual structure can serve as initiation or nucleation sites for conformational transformation during protein−target binding processes (*14*−*17, 19*). In other words, target-induced folding could proceed from these nascent structural regions. This may hold true for rP172 amelogenin with regard to forming supramolecular assemblies within the enamel matrix, where nascent helical or extended β-strand regions (Figure 4) may propagate in response to external stabilization. Second, the location of residual structure and the boundaries between residual structural elements within rP172 amelogenin may have an important impact with regard to molecular recognition, and in particular by amelogenin-specific proteases. It is known that amelogenin is cleaved by specific proteases at specific sites within the primary sequence (*57, 58*). Coincidentally, the common conserved cleavage sites within porcine amelogenin, W45/H46, S148/M149, W161/A162, and H62/H63 (*57, 58*), are positioned at the boundaries of secondary structure regions (Figure 4). Given the unfolded nature of the rP172 amelogenin protein, these cleavage sites are located at the starting points of residual secondary structure and thus would not be obscured by any residual folding whatsoever but would be accessible for protease recognition. Finally, one of the most interesting types of secondary structure, PPII, is well-represented within monomeric rP172 amelogenin [I70−P89 and P102−P145 (Figure 4)], and we note that PPII structure has been linked to protein−protein interaction processes (*49*) and to the stabilization of sequences against conformational transformation (*50*). Thus, the presence of PPII structure within the rP172 amelogenin protein may serve dual roles in terms of promoting molecular recognition and stabilizing unfolded regions against transformation until appropriate conditions are realized. Obviously, additional experimentation will be necessary to firmly establish how these residual structure regions assist in amelogenin functionality.

ACKNOWLEDGMENT

Portions of this work represent contribution 49 from the Laboratory for Chemical Physics (J.S.E.) and were funded by a grant from the Department of Energy (DE-FG02-03ER46099 to J.S.E.). Prof. John Evans is a member of the New York Structural Biology Center, which is a STAR center supported by the New York State Office of Science, Technology, and Academic Research. NMR resources were supported by NIH Grant P41 GM66354. We thank Prof. Uversky for helpful discussions and suggestions regarding amelogenin sequence analysis, Prof. Ralph Lengan for allowing access to the CD instrument, and Kaushik Dutta for assistance in the NMR experimental setup.

SUPPORTING INFORMATION AVAILABLE

rP172 ^{1}H, ^{13}C, and ^{15}N temperature-corrected (283 K) backbone chemical shifts and HNHA-determined $^{3}J_{\mathrm{NHCH}\alpha}$ coupling constants (Table S1), CD spectra of rP172 in UDDWat pH 6.0 and 10 °C and heated to 85 °C and then cooled to 10 °C (Figure S1), and ^{15}N, ^{1}H HSQC spectra of rP172 in UDDW at pH 3.8 (Figures S2 and S3). This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Lowenstam, H. A., and Weiner, S. (1989) *On Biomineralization*, Oxford University Press, New York.
2. Snead, M. L., Zeichner-David, M., Chandra, T., Robson, K. J., Woo, S. L., and Slavkin, H. C. (1983) Construction and identification of mouse amelogenin cDNA clones. *Proc. Natl. Acad. Sci. U.S.A. 80*, 7254–7258.
3. Fincham, A. G., Moradian-Oldak, J., and Simmer, J. P. (1999) The structural biology of the developing dental enamel matrix. *J. Struct. Biol. 126*, 270–299.
4. Fincham, A. G., Moradian-Oldak, J., Diekwisch, T. G., Lyaruu, D. M., Wright, J. T., Bringas, P., Jr., and Slavkin, H. C. (1995) Evidence for amelogenin "nanospheres" as functional components of secretory-stage enamel matrix. *J. Struct. Biol. 115*, 50–59.
5. Moradian-Oldak, J., Du, C., and Falini, G. (2006) On the formation of amelogenin microribbons. *Eur. J. Oral Sci. 114* (Suppl. 1), 289–296, discussion 327−329, 382.

6. Wiedemann-Bidlack, F. B., Beniash, E., Yamakoshi, Y., Simmer, J. P., and Margolis, H. C. (2007) pH triggered self-assembly of native and recombinant amelogenins under physiological pH and temperature in vitro. *J. Struct. Biol 160*, 57–69.
7. Ijima, M., and Moradian-Oldak, J. (2004) Control of octacalcium phosphate and apatite crystal growth by amelogenin matrices. *J. Mater. Chem. 14*, 2189–2199.
8. Beniash, E., Simmer, J. P., and Margolis, H. C. (2005) The effect of recombinant mouse amelogenins on the formation and organization of hydroxyapatite crystals in vitro. *J. Struct. Biol. 149*, 182–190.
9. Delgado, S., Girondot, M., and Sire, J. Y. (2005) Molecular evolution of amelogenin in mammals. *J. Mol. Evol. 60*, 12–30.
10. Gibson, C. W., Yuan, Z. A., Hall, B., Longenecker, G., Chen, E., Thyagarajan, T., Sreenath, T., Wright, J. T., Decker, S., Piddington, R., Harrison, G., and Kulkarni, A. B. (2001) Amelogenin-deficient mice display an amelogenesis imperfecta phenotype. *J. Biol. Chem. 276*, 31871–31875.
11. Paine, M. L., Lei, Y. P., Dickerson, K., and Snead, M. L. (2002) Altered amelogenin self-assembly based on mutations observed in human X-linked amelogenesis imperfecta (AIH1). *J. Biol. Chem. 277*, 17112–17116.
12. Zhu, D. H., Paine, M. L., Luo, W., Bringas, P., and Snead, M. L. (2006) Altering biomineralization by protein design. *J. Biol. Chem. 281*, 21173–21182.
13. Bartlett, J. D., Ganss, B., Goldberg, M., Moradian-Oldak, J., Paine, M. L., Snead, M. L., Wen, X., White, S. N., and Zhou, Y. L. (2006) Protein-protein interactions of the developing enamel matrix. *Curr. Top. Dev. Biol. 74*, 57–115.
14. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol. 5*, 148–155.
15. Dyson, H. J., and Wright, P. E. (1998) Equilibrium NMR studies of unfolded and partially folded proteins. *Nat. Struct. Biol. 7*, 499–503.
16. 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.
17. Ishijima, J., Nagasaki, N., Maeshima, M., and Miyano, M. (2007) RVCaB, a calcium binding protein in radish vacuoles, is predominantly an unstructured protein with a polyproline type II helix. *J. Biochem. 142*, 201–211.
18. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci. 11*, 739–756.
19. Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci. 27*, 527–533.
20. Lakshminarayanan, R. Y., Hegde, B. G., Du, C., Fan, D., and Moradian-Oldak, J. (2009) Analysis of secondary structure and self-assembly by variable temperature circular dichroism and isothermal titration calorimetry. *Proteins: Struct., Funct., Bioinf.* (in press). .
21. Lakshminarayanan, R., Fan, D., Du, C., and Moradian-Oldak, J. (2007) The role of secondary structure in the entropically driven amelogenin self-assembly. *Biophys. J. 93*, 3664–3674.
22. Moradian-Oldak, J., Simmer, J. P., Lau, E. C., Sarte, P. E., Slavkin, H. C., and Fincham, A. G. (1994) Detection of monodisperse aggregates of a recombinant amelogenin by dynamic light scattering. *Biopolymers 34*, 1339–1347.
23. Moradian-Oldak, J., Bouropoulos, N., Wang, L. L., and Gharakhanian, N. (2002) Analysis of self-assembly and apatite binding properties of amelogenin proteins lacking the hydrophilic C-terminal. *Matrix Biol. 21*, 197–205.
24. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins 41*, 415–427.
25. Lim, K. H., Collver, H. H., Le, Y. T. H., Nagchowdhuri, P., and Kenney, J. M. (2007) Characterizations of distinct amyloidogenic conformations of the Aβ(1−40) and (1−42) peptides. *Biochem. Biophys. Res. Commun. 353*, 443–449.
26. 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.
27. Fincham, A. G., and Moradian-Oldak, J. (1993) Amelogenin post-translational modification: Carboxy-terminal processing and the phosphorylation of bovine and porcine "TRAP" and "LRAP" amelogenins. *Biochem. Biophys. Res. Commun. 197*, 248–255.
28. Vuister, G. W., and Bax, A. (1993) Quantitative J correlation: A new approach for measuring homonuclear three-bond J($H^N H^\alpha$) coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc. 115*, 7772–7777.
29. Wang, Y., and Wishart, D. S. (2005) A simple method to adjust inconsistently referenced ^{13}C and ^{15}N chemical shift assignments of proteins. *J. Biomol. NMR 31*, 143–148.
30. 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.
31. Receveur-Brechot, V., Bourhis, J.-M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding. *Proteins 62*, 24–25.
32. Danielsson, J., Jarvet, J., Damberg, P., and Graslund, A. (2005) The Alzheimer β-peptide shows temperature-dependent transitions between left-handed 3_1-helix, β-strand and random coil secondary structures. *FEBS J. 272*, 3938–3949.
33. Bienkiewicz, E. A., Woody, A. Y. M., and Woody, R. W. (2000) Conformation of the RNA polymerase IIC-terminal domain: Circular dichroism of long and short fragments. *J. Mol. Biol. 297*, 119–133.
34. Wustman, B., Santos, R., Zhang, B., and Evans, J. S. (2002) Identification of a "glycine loop"-like coiled structure in the 34-AA Pro, Gly, Met repeat domain of the biomineral-associated protein, PM27. *Biopolymers 65*, 1305–1318.
35. Zhang, B., Wustman, B., Morse, D. E., and Evans, J. S. (2002) Model peptide studies of sequence regions in the elastomeric biomineralization protein, Lustrin A. I. The C-domain consensus -PG-, -NVNCT- motif. *Biopolymers 64*, 358–369.
36. (a) Wishart, D. S., Bigham, 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*, 67–81. (b) Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol. 222*, 311–333.
37. Wishart, D. S., and Case, D. A. (2001) Use of chemical shifts in macromolecular structure determination. *Methods Enzymol. 338*, 3–34.
38. Serrano, L. (1995) Comparison between the ϕ distribution of the amino acids in the protein database and NMR data indicates that amino acids have various ϕ propensities in the random coil conformation. *J. Mol. Biol. 254*, 322–333.
39. 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.
40. Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002) A software tool for the prediction of Xaa-Pro bond conformations in proteins based on ^{13}C chemical shift statistics. *J. Biomol. NMR 34*, 149–154.
41. 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.
42. Hinck, A. P., Eberhardt, E. S., and Markley, J. L. (1993) NMR strategy for determining Xaa-Pro peptide bond configurations in proteins: Mutants of staphylococcal nuclease with altered configuration at proline 117. *Biochemistry 32*, 11810–11818.
43. 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.
44. Andreotti, A. H. (2003) Native state proline isomerization: An intrinsic molecular switch. *Biochemistry 42*, 9515–9524.
45. Nicholson, L. K., and Lu, K. P. (2007) Prolyl cis-trans isomerization as a molecular timer in Crk signalling. *Mol. Cell 25*, 483–486.
46. 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.
47. 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.
48. Wuthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley-Interscience, New York.
49. Cubelis, M. V., Caillez, F., Blundell, T. L., and Lovell, S. C. (2005) Properties of polyproline II, a secondary structure element implicated in protein-protein interactions. *Proteins: Struct., Funct., Bioinf. 58*, 880–892.

50. Darnell, G., Orgel, J. P. R. O., Pahl, R., and Meredith, S. C. (2007) Flanking polyproline sequences inhibit β-sheet structure in polyglutamine segments by inducing PPII-like helix structure. *J. Mol. Biol. 374*, 688–704.
51. Matsushima, N., Izumi, Y., and Aoba, T. (1998) Small angle X-ray scattering and computer aided molecular modeling studies of 20 kDA fragment of porcine amelogenin. Does amelogenin adopt an elongated bundle structure? *J. Biochem. 123*, 150–156.
52. Fincham, A. G., Belcourt, A. B., Termine, J. D., Butler, W. T., and Cothran, W. C. (1981) Dental enamel matrix: Sequences of two amelogenin polypeptides. *Biosci. Rep. 1*, 771–778.
53. Paine, M. L., and Snead, M. L. (1997) Protein interactions during assembly of the enamel extracellular matrix. *J. Bone Miner. Res. 12*, 221–227.
54. Shaw, W. J., Ferris, K., Tarasevich, B., and Larson, J. L. (2008) The structure and orientation of the C-terminus of LRAP. *Biophys. J. 94*, 3247–3257.
55. Gibson, C. W. (2008) The amelogenin "enamel proteins" and cells in the periodontium. *Crit. Rev. Eukaryotic Gene Expression 18*, 345–360.
56. Veis, A. (2003) Amelogenin gene splice products: Potential signaling molecules. *Cell. Mol. Life Sci. 60*, 38–55.
57. Ryu, O. H., Fincham, A. G., Hu, C. C., Zhang, C., Qian, Q., Bartlett, J. D., and Simmer, J. P. (1999) Characterization of recombinant pig enamelysin activity and cleavage of recombinant pig and mouse amelogenins. *J. Dent. Res. 78*, 743–750.
58. Moradian-Oldak, J., Jimenez, I., Maltby, D., and Fincham, A. G. (2001) Controlled proteolysis of amelogenins reveals exposure of both carboxy- and amino-terminal regions. *Biopolymers 58*, 606–616.

BI802175A