



The leucine-rich amelogenin protein (LRAP) is primarily monomeric and unstructured in physiological solution



Barbara J. Tarasevich^{a,*}, John S. Philo^b, Nasib Karl Maluf^b, Susan Krueger^c, Garry W. Buchko^a, Genyao Lin^d, Wendy J. Shaw^a

^a Pacific Northwest National Laboratory, Richland, WA 99354, United States

^b Alliance Protein Laboratories, Inc., San Diego, CA 92121, United States

^c National Institute of Standards and Technology, Gaithersburg, MD 20899, United States

^d WSP Chemicals & Technology, LLC, Leetsdale, PA 15056, United States

ARTICLE INFO

Article history:

Received 14 August 2014

Received in revised form 17 October 2014

Accepted 20 October 2014

Available online 25 October 2014

Keywords:

Amelogenin
LRAP
Enamel
Nanosphere
Biom mineralization

ABSTRACT

Amelogenin proteins are critical to the formation of enamel in teeth and may have roles in controlling growth and regulating microstructures of the intricately woven hydroxyapatite (HAP). Leucine-rich amelogenin protein (LRAP) is a 59-residue splice variant of amelogenin and contains the N- and C-terminal charged regions of the full-length protein thought to control crystal growth. Although the quaternary structure of full-length amelogenin in solution has been well studied and can consist of self-assemblies of monomers called nanospheres, there is limited information on the quaternary structure of LRAP. Here, sedimentation velocity analytical ultracentrifugation (SV) and small angle neutron scattering (SANS) were used to study the tertiary and quaternary structure of LRAP at various pH values, ionic strengths, and concentrations. We found that the monomer is the dominant species of phosphorylated LRAP (LRAP(+P)) over a range of solution conditions (pH 2.7–4.1, pH 4.5–8, 50 mmol/L (mM) to 200 mM NaCl, 0.065–2 mg/mL). The monomer is also the dominant species for unphosphorylated LRAP (LRAP(–P)) at pH 7.4 and for LRAP(+P) in the presence of 2.5 mM calcium at pH 7.4. LRAP aggregates in a narrow pH range near the isoelectric point of pH 4.1. SV and SANS show that the LRAP monomer has a radius of ~2.0 nm and an asymmetric structure, and solution NMR studies indicate that the monomer is largely unstructured. This work provides new insights into the secondary, tertiary, and quaternary structure of LRAP in solution and provides evidence that the monomeric species may be an important functional form of some amelogenins.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The amelogenin proteins are necessary for the formation of tooth enamel, representing 90% of the proteins present in enamel fluid (Termine et al., 1980; Gibson et al., 2001). Humans with mutations in the *AMELX* gene and knock-out mice engineered to have *AMELX* null mutations have highly defective and disorganized enamel structure (Gibson, 2011). Amelogenin proteins are thought to function as a matrix to guide the mineralization of HAP extracellularly because amelogenin nanospheres have been observed along HAP crystallites in immature enamel (Fincham et al., 1995). *In vitro* studies have suggested that amelogenin has roles in initiating nucleation (Tarasevich et al., 2007), controlling growth (Iijima and Moradian-Oldak, 2004), and affecting the spacing of crystallites (Moradian-Oldak et al., 1998a).

Leucine-rich amelogenin protein (LRAP) is a 59-residue splice variant of amelogenin (Fig. 1) (Gibson et al., 1991). Because LRAP appears within the enamel fluid with amelogenin, it has been thought that the protein may have a role in enamel formation (Fincham et al., 1999). *In vivo* studies have shown that LRAP is localized within the extracellular matrix of growing enamel (Gibson et al., 1995) and *in vitro* studies have shown that LRAP can control HAP crystal formation (Le Norcy et al., 2011) suggesting that LRAP, like amelogenin, may have an extracellular matrix function in controlling enamel crystal growth. More recent studies have provided evidence that LRAP may promote enamel growth by acting as a cell signaling molecule, affecting ameloblast differentiation and protein expression. For example, LRAP partially rescued the null amelogenin mouse phenotype (Gibson et al., 2009, 2011), increased enamel growth in tooth explants (Ravindranath et al., 2007) and promoted the differentiation of human enamel organ epithelial cells (Le et al., 2007). LRAP that was overexpressed

* Corresponding author.

in transgenic murine models affected ameloblast differentiation and upregulated amelogenin, MMP-20, and SATB1 proteins (Stahl et al., 2013).

Although LRAP has a role in enamel formation, recent studies have shown that LRAP can also function as a cell signaling protein to promote differentiation of mesenchymal cells (Veis et al., 2000; Warotayanont et al., 2008). LRAP has been found to promote osteogenesis of rat muscle fibroblasts (Veis et al., 2000), cementoblasts (Boabaid et al., 2004), and mesenchymal stem cells (Warotayanont et al., 2009; Wen et al., 2011). A porcine enamel matrix derivative has been shown to have therapeutic applications in promoting the regeneration of cementum and bone in periodontal tissue (Hammarstrom, 1997; Heijl et al., 1997; Sculean et al., 2008) and LRAP has been shown to be the main factor within enamel matrix derivatives to promote osteogenesis (Amin et al., 2012).

In spite of the importance of LRAP as an enamel former, osteogenic protein, or regenerative agent, very little is known about its tertiary and quaternary structure. It will be difficult to fully understand LRAP's biological function or therapeutic potential without a better grasp of its structure. The dominant quaternary structure of full-length amelogenin in solution is the “nanosphere,” aggregates of amelogenin monomers that are 20–60 nm in diameter (Moradian-Oldak, 2001). The nanospheres are considered to be self-assemblies because they are believed to have a hierarchical structure consisting of highly ordered oligomers (Fang et al., 2011). Nanospheres are present in solutions with pH values greater than 6.0 and salt concentrations ranging from 50 mmol/L (mM) to 200 mM (Moradian-Oldak et al., 1994, 1998). Oligomers have also been detected in solutions at pH 5.5 ranging in size from dimers to octamers depending on the protein concentration (Bromley et al., 2011).

Recently we found that phosphorylated LRAP exists as a monomer in 150 mM NaCl, pH 7.4 solutions (Tarasevich et al., 2013). Since this quaternary form is in contrast to the nanospheres observed for full-length amelogenin at pH 7.4, we wanted to further investigate the quaternary structure of LRAP over a range of solution conditions including those found in developing enamel (Sasaki et al., 1991; Robinson et al., 1995). LRAP was studied in solutions ranging from pH 3.0 to 8.0, 50 to 200 mM NaCl, and concentrations from 0.065 to 2 mg/mL. SV was used to determine the distribution of protein species present in the solutions (Lebowitz et al., 2002). SV has several advantages over other analysis techniques, such as dynamic light scattering (DLS), because it uses a centrifugal field to separate species from monomers to large aggregates based on the species' masses and shapes, whereas DLS relies on a mathematical separation. Although DLS is well suited for the measurement of sizes in monodisperse systems, it is not as accurate in characterizing polydisperse systems. Light scattering intensity has a size dependence of radius to the sixth power (r^6) so that large structures can dominate scattering and prevent the detection of smaller structures (Filipe et al., 2010). Also, DLS was not able to resolve particle populations that differed in size by a factor of 2 or less (Filipe et al., 2010). In contrast, SV can directly quantify the proportion of different species in solution on a weight basis, since SV typically uses UV absorbance to monitor sample separation. SV

and SANS were also used to determine the size and shape of the LRAP monomers and solution NMR was used to study LRAP secondary structure.

2. Materials and methods¹

2.1. Materials

Deuterium oxide (99.9%) was purchased from Cambridge Isotopes Laboratories and used as received. All other chemicals were reagent grade and obtained from Sigma–Aldrich.

2.2. Solid-phase LRAP(+P) and LRAP(–P) synthesis

Murine LRAP with phosphoserine (LRAP(+P)) and normal serine (LRAP(–P)) at position 16 (Fig. 1) was synthesized using solid-phase methods by the Protein Chemistry Technology Center, University of Texas (Dallas, TX) for SV, SANS, and zeta potential experiments. Each sample was purified by reverse phase HPLC using buffer A, 0.1% trifluoroacetic acid in water and buffer B, 0.1% trifluoroacetic acid in acetonitrile. LRAP eluted at 54% B. Mass spectroscopy was used to characterize the purity and molecular weight of the proteins. After purification, proteins were lyophilized for storage until ready for use.

2.3. Recombinant LRAP(–P) (rLRAP(–P)) synthesis

To obtain uniformly ¹³C- and ¹⁵N-labeled LRAP samples for solution NMR studies it was necessary to use recombinant methods. The protein, rLRAP(–P) containing a 12-residue (MRGSHHHHHGS-) N-terminal tag, was prepared for the studies conducted in acetic acid at pH 2.8 as described previously (Buchko et al., 2010). For the studies conducted in SCP solution at pH 7.4, rLRAP(–P) was prepared with only a 4-residue (GPGS-) N-terminal tag as described in the Supplementary section.

2.4. LRAP solution formation

A stock solution of LRAP was dissolved in Millipore purified water at 5–10 mg/ml and stored at 4 °C overnight. The solution was centrifuged at 11,000 rpm for 30 min and the upper fraction decanted. The concentration of the stock solution was determined by measurement of the UV absorbance at 275 nm using an extinction coefficient of 15,470 ($M^{-1} cm^{-1}$). Saturated calcium phosphate (SCP) solutions were prepared containing various NaCl concentrations (50–300 mM) and pH values (5.0–8.0). They were prepared by adding HAP powder to the NaCl solution, adjusting the pH, stirring for several days, and filtering out any particles as described previously (Shaw et al., 2004). A SCP solution is being used because it has a small amount of dissolved calcium and phosphate in the solution making it more relevant to the *in vivo* enamel forming environment. The amounts of calcium and phosphate do not exceed the saturation limit so there is no calcium phosphate precipitation. The LRAP stock solution was diluted into the various SCP solutions at the appropriate pH to obtain solutions containing 50, 150, and 200 mM NaCl. The solution pH values were initially ~pH 3.0 (due to formic acid in the lyophilizing solution) and were readjusted to values ranging from pH 3.0 to pH 8.0 using dilute

```
MPLPPHPGSP GYINLSYEV L TPLK WYQSMI RQPYPSYGYE PMGGWLHHQI
IPVLSQQHPP SHTLQPHHLL FVVPAAQQPVA PQQPMMPVPG HHSMTPTQHH
QPNIPPSAQQ PFOQPFQQA IPPQSHQPMQ PQSPLHPMQ LAPQPLPPL
FSMQPLSPIL PELPLEAWPA TDKTKREEVD
```

Fig. 1. Primary amino acid sequence of murine amelogenin with the basic and acidic residues colored blue and red, respectively. The splice variant LRAP is composed of the N-terminal 33 and C-terminal 26 residues, shaded yellow. Both full-length amelogenin and LRAP are post-translationally modified by side chain phosphorylation of S16, colored green. The serine at position 16 was either phosphorylated (LRAP(+P)) or non-phosphorylated (LRAP(–P)).

¹ Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

KOH and HCl. The LRAP solution pH values were stable and did not change over time periods up to 1 month. The final concentrations of the LRAP solutions were 0.065–2 mg/ml.

2.5. Zeta potential

The zeta potential of LRAP solutions at concentrations of 0.5 mg/mL, 50 and 150 mM NaCl SCP, and various pH values from 3.0 to 8.0 were determined using a ZetaPALS instrument (Brookhaven Instruments, Holtsville, NY). The protein electrophoretic mobility was measured by detecting the phase shift in the scattered light as the protein moves under an applied field. The protein solution was placed into a cuvette and an Uzgiris cell electrode was inserted into the solution. The Uzgiris cell suppressed electro-osmosis effects. The palladium electrodes were conditioned by 10 runs (30 cycles/run) of a 1 M NaCl solution with both voltage and frequency set to auto. The resulting PdCl₂ passivating layer reduced chlorine and protein interactions with the charged electrode. The electrode was operated at 2.5 V and 2.5 Hz for the protein solutions and the data was collected over five runs at 10–20 cycles per run.

2.6. Sedimentation velocity (SV)

Samples were analyzed within 2–3 days after sample preparation and were placed into an AN-60Ti analytical rotor, loaded into a Beckman-Coulter ProteomeLab XL-I analytical ultracentrifuge at 20 °C, and scanned at 280 nm or 230 nm as described previously (Tarasevich et al., 2013). The data were analyzed using the *c(s)* method and SEDFIT analysis program (version 11.3) developed previously (Schuck, 2000). The raw data scans (~37,000 data points) were directly fitted to derive the distribution of sedimentation coefficients, while modeling the influence of diffusion on the data in order to enhance the resolution. This method assigns a diffusion coefficient to each value of sedimentation coefficient based on an assumption that all species have the same overall hydrodynamic shape. That hydrodynamic shape is defined by the *ff/f₀* ratio, where *f* is the frictional coefficient of the macromolecule and *f₀* is the frictional coefficient of an anhydrous sphere with the same volume as the macromolecule. The *ff/f₀* values were varied to find the best overall fit of the data for the entire sample. Maximum entropy regularization probabilities of 0.683 (1 σ) were used and time-independent noise was removed.

The partial specific volume (\bar{v}) for LRAP(+P) and LRAP(–P) at 20 °C were calculated as 0.7435 and 0.7476 ml/g, respectively, using the program SEDNTERP (version 1.09) (Laue et al., 1992). Solvent densities and viscosities at 20 °C for the SCP buffer were calculated using SEDNTERP as 1.00442 g/ml and 1.0165 cp respectively. SEDNTERP was also used to convert raw sedimentation coefficients to standardized *s*_{20,w} values and to calculate *ff/f₀* ratios from the measured sedimentation coefficients.

2.7. Small angle neutron scattering (SANS)

LRAP(+P) solutions were studied by SANS to determine the size and shape of the LRAP monomers. LRAP(+P) (1 mg/mL) was prepared in 150 mM NaCl SCP (pH 7.4) or 2% acetic acid (pH 2.7) in 99.9% D₂O and was measured in a quartz cuvette with a 2 mm path length on the NG7 30 meter SANS instrument (Glinka et al., 1998) at the NIST Center for Neutron Research (NCNR) in Gaithersburg, MD. The 2% acetic acid sample was also measured at a LRAP(+P) concentration of 2 mg/mL. Preliminary experiments showed that the SANS data for pH 7.4 samples was affected by the presence of oligomers derived from aggregates of LRAP(+P) formed at the isoelectric point. At oligomer concentrations of 5% or greater, the oligomer species dominated the scattering and it was not possible to obtain both a size distribution from the fit and extract size and

shape data for the monomer. Methods were developed to reduce the concentration of oligomers in the solutions to 1–2% to allow characterization of the monomers. This was achieved by reducing the protein concentration and adding LRAP stock solution to buffer containing enough base to obtain a final pH of 7.4, avoiding the pH range where LRAP has a low charge and aggregates (pH 4.2–4.5).

The neutron wavelength, λ , was 6 Å, with a wavelength spread, $\Delta\lambda/\lambda$, of 0.15. Scattered neutrons were detected with a 64 × 64 cm two-dimensional position-sensitive detector with 128 × 128 pixels at a resolution of 0.5 cm/pixel. Sample-to-detector distances of 5 m and 1.5 m were used. The data were reduced using the IGOR program with SANS macro routines developed at the NCNR (Kline, 2006). Raw counts were normalized to a common monitor count and corrected for empty cell counts, ambient room background counts, and non-uniform detector response. Data were placed on an absolute scale by normalizing the scattered intensity to the incident beam flux. Next, the data were radially averaged to produce scattered intensity, *I(q)*, versus *q* curves, where $q = 4\pi\sin(\theta)/\lambda$ and 2θ is the scattering angle. The scattered intensities were then further corrected for buffer scattering and incoherent scattering from hydrogen in the samples. The *q*-range covered by the data after background subtraction was $0.012 \text{ \AA}^{-1} \leq q \leq 0.3 \text{ \AA}^{-1}$ for the 2 mg/mL samples and $0.03 \text{ \AA}^{-1} \leq q \leq 0.3 \text{ \AA}^{-1}$ for the 1 mg/mL samples.

Initial data analysis was performed using the Guinier approximation,

$$I(q) = I(0) \exp\left(-q^2 R_g^2/3\right), \quad (1)$$

on the low-*q* portions of the data to obtain initial values for the radius of gyration, *R_g*, and the forward scattering intensity, *I(0)*, of the samples. This analysis is valid only in the region where $qR_g \leq 1$. Standard Kratky analysis was performed on each data set by plotting *I(q)·q²* versus *q* to get a qualitative idea of the degree of protein folding. Both the Guinier and Kratky analyses were performed using the NCNR IGOR SANS macro routines (Kline, 2006).

GNOM was used to determine the distance distribution functions, *P(r)*, versus *r* (Semenyuk and Svergun, 1991). The value of the maximum diameter of the particle, *D_{max}*, was determined empirically by examining the quality of the fit to the experimental data for a range of *D_{max}* values. Since the SANS data were obtained on an absolute scale, the molecular weight, *M_w*, of the proteins was estimated from the forward scattering intensity, *I(0)*, using the relation:

$$I(0) = n(\Delta\rho V)^2, \quad (2)$$

where $\Delta\rho = \rho - \rho_s$ is the contrast, or the difference between the scattering length density of the molecule (ρ) and the solvent (ρ_s), *n* is the number density of molecules, and *V* is the molecular volume. The number density can be written as $n = cN_A/M_w$, where *c* is the concentration, and *N_A* is Avogadro's number. The volume can be written as $V = M_w/(N_A d)$, where *d* is the mass density. Now, Eq. (2) can be rewritten as:

$$\frac{I(0)}{c} = \frac{(\Delta\rho)^2}{N_A d^2} M_w. \quad (3)$$

The only unknown parameter in Eq. (3) is *M_w*, since all other parameters can be measured or calculated. *I(0)* values were taken from the GNOM analysis of the data since this analysis uses all of the data rather than a limited number of points in the Guinier region. The concentration can be directly measured during sample preparation and $\Delta\rho$ can be calculated from the chemical composition of the sample and solvent. The mass density, *d*, is taken as the inverse of the partial specific volume. It is important to note that *I(0)* must be on an absolute scale, usually in cm^{–1}, in order to obtain accurate *M_w* values from Eqs. (2) or (3).

2.8. Molecular modeling

Model SANS curves were calculated for the 10 lowest energy Rosetta model structures of LRAP(+P) derived previously (Tarasevich et al., 2013) using the program XTAL2SAS (Heidorn and Trewhella, 1988; Krueger et al., 1998). The model SANS curves were compared to the SANS data from 2 mg/mL LRAP(+P) in 2% acetic acid and scored for quality using the χ^2 equation:

$$\chi^2 = \frac{1}{(N-1)} \frac{\sum_q (I_{\text{exp}}(q) - I_{\text{calc}}(q))^2}{\sigma_{\text{exp}}(q)^2} \quad (4)$$

where $I_{\text{exp}}(q)$ is the experimentally determined SANS intensity curve, $I_{\text{calc}}(q)$ is the calculated intensity curve from the model structure, and $\sigma_{\text{exp}}(q)$ is the q -dependent variance. The sum was taken over 60 data points ($N = 60$). The lowest energy structure with the model SANS curve that best fit the data was used as a starting point for further structure modeling.

The SASSIE program (Curtis et al., 2012) was used to generate an ensemble of structures from this best-fit starting structure for comparison to the SANS data. A variety of structures were generated by randomly varying backbone dihedral angles within the regions of the protein that are disordered in the Rosetta model structures, i.e., residues 1–22, residues 33–43, residues 48–49 and residues 57–59. After each randomly-chosen angle was rotated by a given value, the final value was checked to determine if it was energetically probable, based on an energy function using the Charmm-22 all-atom protein force-field parameters (MacKerell et al., 1998). Each new configuration was checked for overlap of α -carbon atoms using a distance criterion of 3 Å. If both checks were met, the new structure was accepted and R_g was calculated. All structures were energy minimized using the program NAMD (Phillips et al., 2005) prior to the calculation of model SANS curves.

Examination of a plot of χ^2 versus R_g provides an idea of how well the individual structures in the ensemble fit the data. The best (lowest χ^2) and worst (highest χ^2) fit model SANS curves are noted, along with the average model SANS curve from the entire ensemble of accepted structures. These curves were plotted along with the experimental SANS data to aid in the visualization of the quality of the fits to the data.

2.9. Nuclear magnetic resonance (NMR) spectroscopy

Varian Inova-500 and -750 spectrometers equipped with an HCN-probe and pulse field gradients were used to collect NMR data at 20 °C on the double-labeled (^{13}C -, ^{15}N -) rLRAP(-P) sample (2 mg/ml) in 2% acetic acid (pH 2.8) and in 150 mM NaCl SCP solutions (pH 7.4). Three-dimensional HNCACB and CCC-TOCSY data were collected to assign the amide cross peaks in the two-dimensional ^1H - ^{15}N HSQC spectrum and to acquire $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts. All NMR data were processed using Felix2007 (MSI, San Diego, CA) software and analyzed with the program Sparky (v3.115) (Goddard). The ^1H , ^{13}C , and ^{15}N chemical shifts were referenced using indirect methods (Wishart et al., 1995). The random coil values used in the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift plots were from CNS (cns_solve_1.1).

3. Results

3.1. Zeta potential of LRAP(+P) as a function of pH

Fig. 2 shows the zeta potential of LRAP(+P) solutions at 0.5 mg/mL in SCP solution as a function of pH and ionic strength. The zeta potential is positive at low pH and negative at high pH with an

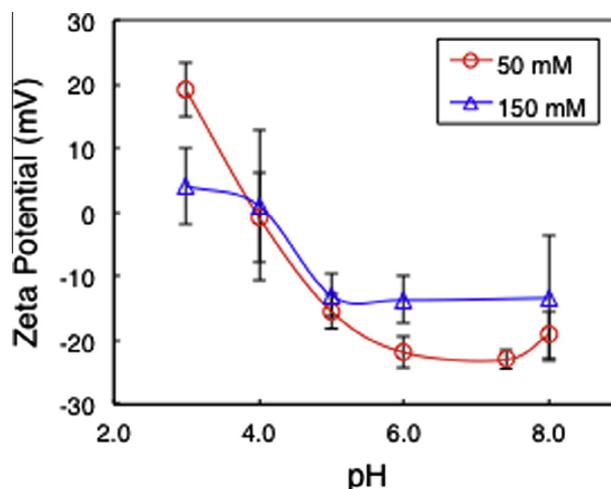


Fig. 2. Zeta potential of LRAP (+P) in SCP solution at 50 and 150 mM NaCl as a function of pH. Data points are connected by smoothed lines.

isoelectric point around pH 4.1. The zeta potential decreases with increasing ionic strength from 50 mM to 150 mM NaCl.

3.2. SV studies of quaternary structures

Standardized sedimentation coefficient ($s_{20,w}$) distributions were obtained for LRAP(+P) in SCP solution at 150 mM NaCl, pH 7.4 as a function of LRAP concentration. The solutions were formed at 2 mg/mL LRAP(+P) and then diluted in SCP solution to 1.0 and 0.5 mg/mL. The species distribution for the 2 mg/mL solution is shown in Fig. 3, distributions for 0.5 and 1.0 mg/mL LRAP(+P) are shown in Fig. S1, and the sedimentation coefficients ($s_{20,w}$) and weight % species for all concentrations are summarized in Table 1. The size distributions are dominated by one major peak for all three concentrations. The sedimentation coefficient, together with the best-fit frictional coefficient ratio relative to an anhydrous sphere (f/f_0 ratio), imply this major peak has a mass of ~ 6.3 kDa, consistent with a monomeric species. The assignment of the major component to a monomer was also confirmed by time derivative

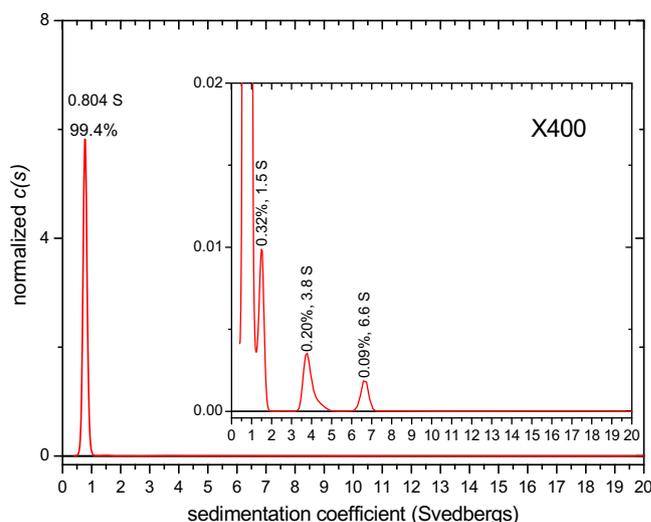


Fig. 3. Normalized sedimentation coefficient distribution for LRAP(+P) in 150 mM NaCl, pH 7.4 SCP solution at 2 mg/mL with a 400 \times expansion in the inset. Peaks are labeled with % amount and sedimentation coefficient, $s_{20,w}$ value in svedbergs (S). The main peak was determined to represent a monomer and the $s_{20,w}$ value of the monomer peak is the average over the peak.

analysis (Philo, 2006). Repetitions of samples showed that the sedimentation coefficient values for the monomers are reproducible with errors of 1.4% and 2.4% for the 1 and 2 mg/mL solutions, respectively (Table S1).

The 0.5 mg/mL solution is 100% monomer and the solutions at higher concentration have a very small percentage of larger species (0.2% and 0.6% for the 1.0 mg/mL and 2.0 mg/mL solutions, respectively). The peaks at 1.5 S, 3.6 S, and 6.6 S for the 2.0 mg/mL solution and 5.3 S, 10.5 S, and 13.5 S for the 1.0 mg/mL solution are suggestive of oligomers with a range of sizes. The oligomer species are reported as two classes in Table 1, small oligomers from 1.1 S to 1.6 S, and oligomers larger than 1.6 S. The peak positions and concentration of the oligomers vary between repetitions of the same sample as shown in Fig. S2. Also, there are differences in oligomer concentration and distribution with changes in the sample preparation method. As discussed in detail later, we believe that the oligomers are derived from larger aggregates formed at the isoelectric point of LRAP. The oligomers, therefore, are metastable, nonequilibrium species originating from the sample preparation method. The peak positions for the smallest minor components are in the range of 1.1–1.6 S, consistent with species in the size range of dimers to tetramers.

Although the sedimentation coefficient for pure monomer should decrease with increasing protein concentration by at least 0.9% per mg/mL due to excluded volume effects (Laue et al., 1992), there is no significant change in sedimentation coefficient with increasing LRAP(+P) protein concentration (Tables 1 and S1). The sedimentation coefficient values, therefore, are larger than expected at the higher protein concentrations. This suggests that the protein is exhibiting rapidly reversible, self-association reactions on a time scale too short to result in any resolved oligomer peaks. The peak we have been describing as a “monomer” at concentrations above 0.5 mg/mL, therefore, represents a dynamic mixture of monomers and a few percent of short-lived dimers or other larger species. In summary, there are two kinds of structures in the LRAP solutions: a large concentration of monomers in rapid, reversible association with other monomers and a small concentration of oligomers that are metastable.

In addition to the pH 7.4, 150 mM NaCl SCP conditions, we also studied LRAP(+P) structure using SV under a range of solution conditions from 50 to 200 mM NaCl and pH values of 5.8–7.4. The results are summarized in Table 2. The data show that the monomer is the dominant species over all of the solution conditions and pH values studied. The monomer percentage was as low as 86.4% for the 50 mM NaCl, pH 5.8 LRAP(+P) solution and as high as 97.9% for the 200 mM NaCl, pH 8.0 LRAP(+P) solutions.

Although monomers are the dominant species over most of the pH range, we observed large structures by DLS (~1400 nm diameter) and visible cloudiness in 1 mg/mL solutions in the pH range 4.2–4.4. Since our LRAP solutions are below the solubility limit for this pH range (Tan et al., 1998), we believe the structures are LRAP(+P) aggregates formed at low charge near the isoelectric point of LRAP(+P) as determined by the zeta potential measurements shown in Fig. 2. Previous dynamic light scattering (DLS) studies also showed aggregation of 2 mg/mL LRAP in this pH range (Le Norcy et al., 2011). As the pH is adjusted above pH 4.4, the vis-

ible cloudiness disappears, the solutions become clear, and monomers become the dominant species. Above the isoelectric point, therefore, LRAP develops a negative charge which promotes dissolution of the aggregates to monomers and a small fraction of oligomers as small as dimers and trimers. Since the oligomers do not dissolve completely to the equilibrium monomeric species in the time course of the experiments, they represent metastable, non-equilibrium species. The % oligomeric species are highest at pH 5.8 and there is a general trend that the % oligomers decrease with increasing salt concentration from 50 mM to 200 mM NaCl. These trends may represent variations in the degree of dissolution of the aggregates formed near the isoelectric point.

Previous transmission electron microscopy (TEM) and atomic force microscopy (AFM) studies have observed LRAP “nanospheres” adsorbed onto surfaces (Le Norcy et al., 2011; Habelitz et al., 2006; Tarasevich et al., 2010). Our SV studies have found that LRAP solutions contain primarily monomers and up to 14% oligomers. Our research, therefore, suggests that the “nanospheres” seen previously on surfaces may be the solution structures we are calling “nonequilibrium oligomers.” The monomers we observe as the majority species may have also adsorbed onto the surfaces previously studied but may have been difficult to detect experimentally.

The monomer is also the dominant species in solutions containing LRAP(+P) in 2.5 mM CaCl₂, pH 7.4 solutions and LRAP(–P) in 150 mM NaCl, pH 7.4 SCP solutions as shown in Table S2. Previous small angle X-ray scattering (SAXS) studies determined that LRAP(+P) in similar calcium-containing solutions (2 mg/mL, 2.5 mM CaCl₂, pH 7.4) had a radius of gyration, R_g , of 4.4 nm (Le Norcy et al., 2012). This result was interpreted to suggest that LRAP formed aggregates in the presence of calcium. Since we found 94% monomers and 6% oligomers in our samples at a similar solution condition (Table S2), another way to interpret the SAXS data is that the 4.4 nm diameter represents the average size over the monomers and oligomers present in solution. SAXS is a scattering technique that measures the average size over the intensity weighted distribution of species in contrast to SV which determines a species distribution. We found similar large average sizes of LRAP in our SANS scattering studies if we had 5–10% oligomers in our solutions. The presence of oligomers mixed with monomers, therefore, can shift up the average size of species studied by scattering techniques such as SAXS and SANS. If we were careful about making our solutions and reduced the % oligomers down to 1%, we obtained SANS sizes that were representative of the monomer (~2 nm R_g).

3.3. SV studies of monomer tertiary structure

Since the monomer is the dominant quaternary structure of LRAP observed over a range of protein concentrations, pH values, salt concentrations, and phosphorylation state, we decided to further study the tertiary structure of the LRAP monomer by determining its hydrodynamic radius (R_h) and f/f_0 ratio, both of which were determined from the observed sedimentation coefficient and the known monomer mass. The f/f_0 ratio is the frictional coefficient of the macromolecule relative to the frictional coefficient of

Table 1
Sedimentation coefficient ($s_{20,w}$) and % species for LRAP(+P) in 150 mM NaCl, pH 7.4 SCP solution as a function of LRAP concentration.

[LRAP(+P)] (mg/mL)	[NaCl] (mM)	pH	Monomer ^a $s_{20,w}$ (Svedbergs)	Monomer ^a w%	~1.1–1.6 S peaks %	>1.6 S peaks %
0.5	150	7.4	0.781	100	0	0
1	150	7.4	0.782	99.8	0	0.2
2	150	7.4	0.804	99.4	0.3	0.3

^a The monomer peak at these high concentrations represent monomers in weak, rapid reversible association with monomers or other species.

Table 2
Sedimentation velocity data for LRAP(+P) in SCP solutions at various NaCl concentrations and pH values.

[LRAP(+P)] (mg/mL)	[NaCl] (mM)	pH	Monomer ^a %	~1.1–1.6 S peaks %	>1.6 S peaks %
1.0	50	5.8	86.4	7.0	6.6
0.78	50	7.4	93.6	3.3	3.1
0.78	50	8.0	92.3	4.7	3.0
1.0	150	5.8	90.9	7.1	2.0
1.0	150	7.4	93.9	5.8	0.3
0.92	150	8.0	92.6	7.2	0.2
1.0	200	5.8	94.5	0	5.5
1.0	200	7.4	97.7	0	2.3
0.92	200	8.0	97.9	0	2.1

^a The monomer peak at these high concentrations represent monomers in weak, rapid reversible association with monomers or other species.

an anhydrous sphere and is a function of the shape and hydration of the species. In the absence of hydration, an fff_0 ratio of 1 indicates a spherical shape, while an fff_0 ratio greater than 1 suggests an asymmetric shape. We obtained sedimentation coefficient distributions and fff_0 ratios of LRAP as a function of solution pH at solution concentrations of 0.065 mg/mL. A low protein concentration was used to decrease the extent of self-association of LRAP monomers which would affect the measured $s_{20,w}$ values. The fff_0 ratio was then calculated from the measured sedimentation coefficient of the monomer peak and the known monomer mass.

Table 3 is a summary of sedimentation coefficient values, R_h values, and fff_0 ratios for the monomer ordered by pH and Table S3 shows a summary of percent species for the various solution conditions. Repetitions of LRAP(+P) in 150 mM NaCl, pH 7.4 SCP samples resulted in errors of ~0.5% for the sedimentation coefficient, R_h , and fff_0 ratio values. The hydrodynamic radius, R_h , of the LRAP(+P) aggregate at pH 4.2 was also determined by DLS. Table 1 and the plot in Fig. 4a show that the solutions contain ~2 nm radius monomers over the entire pH range except near the isoelectric point where 130 nm aggregates are formed. Fig. 4b shows the fff_0 ratio plotted as a function of solution pH. The fff_0 ratio for SCP solutions at 50 mM NaCl is highest at pH 3 (1.64) and decreases slightly with increasing pH. LRAP has fff_0 ratios ranging from 1.52 to 1.64 over the solution conditions used, indicating that the protein has a more extended or flexible structure than globular proteins (Schuck, 2000).

Table 3 shows that the R_h also decreases slightly for LRAP(+P) at 50 mM NaCl with increasing pH, from 2.09 at pH 3 to 2.02 at pH 8. Increasing the NaCl concentration to 150 mM at pH 7.4 results in a smaller R_h and fff_0 ratio suggesting a more compact structure. LRAP(+P) and LRAP(–P) have the same size and shape in 150 mM, pH 7.4 SCP solutions suggesting that phosphorylation of S16 has little effect on the protein's tertiary structure at that pH. LRAP(+P) solutions containing 2.5 mM CaCl_2 at pH 7.4 have the

Table 3
Sedimentation velocity results including hydrodynamic radius (R_h) and fff_0 ratios for the LRAP monomer in dilute solutions (0.065 mg/mL) as a function of pH, and DLS results for the hydrodynamic radius of LRAP aggregates at pH 4.2.

LRAP	Buffer type	[NaCl] (mM)	[CaCl ₂] (mM)	pH	Monomer $s_{20,w}$ (Svedbergs)	R_h (nm)	fff_0 ratio
+P	SCP	50	–	3.0	0.740	2.09	1.64
+P	SCP	50	–	4.0	0.749	2.06	1.62
+P	SCP	50	–	4.2	130 ^a	–	–
+P	SCP	50	–	5.0	0.760	2.03	1.60
+P	SCP	50	–	6.0	0.760	2.03	1.60
+P	SCP	50	–	7.4	0.761	2.03	1.60
+P	SCP	50	–	8.0	0.765	2.02	1.59
+P	SCP	150	–	7.4	0.781	1.98	1.56
–P	SCP	150	–	7.4	0.758	1.98	1.56
+P	CaCl_2	3.1	2.5	7.4	0.799	1.93	1.52

^a R_h of aggregates determined by DLS.

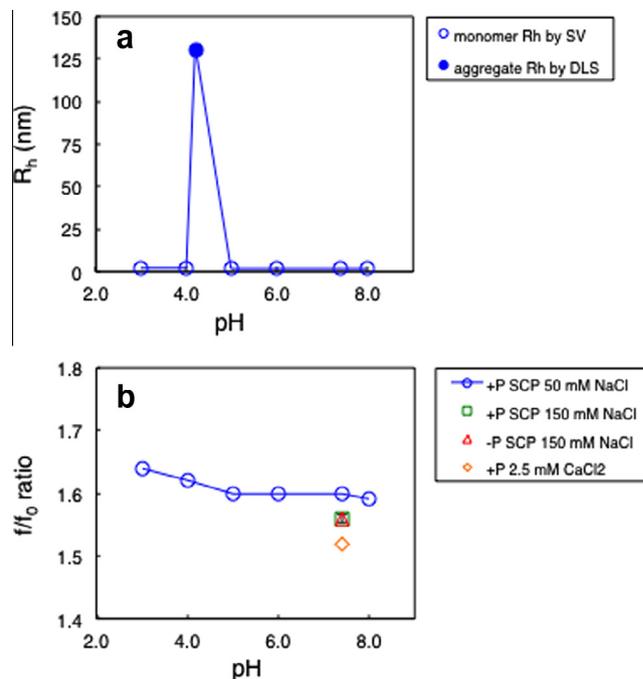


Fig. 4. (a) R_h for LRAP(+P) species in 0.065 mg/ml 50 mM NaCl SCP solutions including R_h of monomers determined by SV (open blue circles) and R_h of aggregates at pH 4.2 determined by DLS (filled blue circle), (b) SV-determined fff_0 ratios for monomers in 0.065 mg/mL LRAP solutions including LRAP(+P) in 50 mM NaCl SCP (blue circles), LRAP(+P) in 150 mM NaCl SCP at pH 7.4 (green square), LRAP(–P) in 150 mM NaCl SCP at pH 7.4 (red triangle), and LRAP(+P) in 2.5 mM CaCl_2 at pH 7.4 (orange diamond). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

smallest R_h (1.93 nm) and smallest fff_0 ratio (1.52), indicating that the monomer collapses in the presence of calcium. A previous small angle X-ray scattering (SAXS) study also showed that LRAP is more compact in the presence of calcium (Le Norcy et al., 2012).

3.4. SANS studies of tertiary structure

The $I(q)$ versus q SANS curves are shown in Fig. 5 and the corresponding Guinier, Kratky, and $P(r)$ versus r plots are provided in the Supplementary section (Figs. S3–S5). Fig. 5a shows the SANS data for 1 mg/mL LRAP(+P) in 2% acetic acid (pH 2.7) and in 150 mM NaCl, SCP solution (pH 7.4) where the two curves are identical within experimental error. Fig. 5b shows the SANS data for LRAP(+P) in 2% acetic acid at 1 and 2 mg/mL where the curves are also identical within experimental error. Table 4 lists the radius of gyration, R_g , and $I(0)$ values determined for these samples from

the SANS data and those calculated from the LRAP amino acid sequence.

Since all of the SANS curves are essentially identical, we chose to perform structural modeling on the 2 mg/mL LRAP(+P), 2% acetic acid (pH 2.7) data set because it contained the best signal-to-noise and the widest q -range. The lowest energy Rosetta model structure that best fit the SANS data had a χ^2 value of 1.6. Using this Rosetta model structure as the starting point, an ensemble of 9427 accepted, energy-minimized structures was generated with SASSIE. Model SANS curves were calculated for each of the structures in the ensemble and compared to the data. Fig. 6a shows the original 2 mg/mL 2% acetic acid, pH 2.7 SANS data along with model SANS traces for the average and best-fit SASSIE structures. Both model traces lie within the experimental limits of the original data showing that the SASSIE analyses generated structures that fit the SANS data well. Fig. 6b is a plot of χ^2 versus R_g that includes the SASSIE-derived structures along with the 10 lowest energy Rosetta structures. Clearly, the SASSIE ensemble provides a broader distribution of structures that fit the SANS data than the 10 best Rosetta model structures. The distribution of SASSIE structures indicates that LRAP(+P) can adopt a range of conformations dominated by extended structures with R_g of 1.7–2.3 nm.

3.5. NMR studies of the secondary structure of rLRAP(–P)

Previous studies of the secondary structure of full-length amelogenin have been performed using NMR at low pH (~ 3 – 4) where amelogenin exists as a monomer (Buchko et al., 2008; Delak et al., 2009). It is not possible to obtain structural information of the amelogenin monomer within the oligomer or nanosphere at physiological conditions (pH 5.8–8.0) using traditional NMR methods because of severe spectral line broadening. Our discovery that LRAP is a monomer under most solution conditions provides an opportunity to study the secondary structure of an amelogenin under physiological conditions using NMR. We were limited to studying LRAP with unphosphorylated S16 (LRAP(–P)) because NMR studies required recombinant methods to uniformly label LRAP with ^{15}N and ^{13}C . It is not possible to recombinantly prepare LRAP and enzymatically phosphorylate only one serine residue at S16. Attempts were made to obtain HSQC spectra on unlabeled

LRAP(+P) but the concentration of naturally abundant ^{15}N and ^{13}C was not high enough to give good spectra.

Regions of secondary structure in a protein may be identified from the chemical shifts of the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ side chain atoms of amino acid residues. Relative to random coil chemical shift values ($\Delta\delta^{13}\text{C} = \delta_{\text{Random coil}} - \delta_{\text{observed}}$), these carbon chemical shifts are associated with α -helical (negative $^{13}\text{C}^\alpha$, positive $^{13}\text{C}^\beta$) or β -strand (positive $^{13}\text{C}^\alpha$, negative $^{13}\text{C}^\beta$) secondary structure. The deviations of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ for rLRAP(–P) in 150 mM NaCl, pH 7.4 SCP solution are shown in Fig. 7. Most of the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts for rLRAP(–P) do not deviate significantly from random coil values, with most differences less than 1 ppm. Furthermore, there are no consecutive runs for four or more residues of corresponding positive/negative or negative/positive deviations. Similar observations are made for the deviations of the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ of rLRAP(–P) in 2% acetic acid (pH 2.8) (Fig. S6). The solution NMR data for rLRAP(–P), therefore, are consistent with random coil structures in the C-terminal and N-terminal regions. It was not possible to obtain $^{13}\text{C}^\beta$ data in the K24–I30 region where a canonical α -helix has been observed by solid state NMR (SSNMR) and Rosetta simulations (Masica et al., 2011). The NMR data suggest a largely disordered structure for rLRAP(–P), similar to the disordered, extended conformation reported for full-length porcine amelogenin at pH 3.8 in solution (Delak et al., 2009). Circular dichroism (CD) spectra obtained for the phosphorylated LRAP(+P) in SCP at pH 7.4 and in solutions with 3 mM CaCl_2 (Fig. S7) show that LRAP(+P) also has a largely disordered secondary structure.

4. Discussion

4.1. LRAP quaternary structure

Although numerous studies have demonstrated that full-length amelogenin can self-assemble into quaternary structures called nanospheres (Fincham et al., 1995; Moradian-Oldak et al., 1998b), there has been little research on the quaternary structures of LRAP. Our work shows that LRAP(+P) is primarily a monomer over a wide range of protein concentrations, salt concentrations, pH values, and in the presence of calcium. The only solution condition where the monomer is not the dominant species is near the

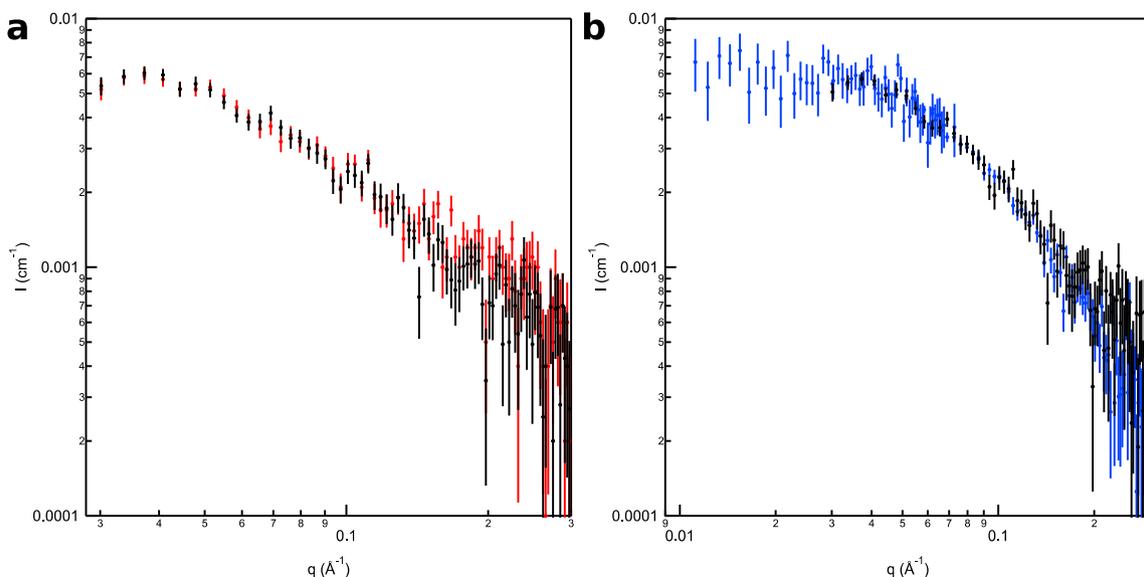


Fig. 5. SANS $I(q)$ versus q data for (a) 1 mg/mL LRAP(+P) in 2% acetic acid (pH 2.7) (black) and 150 mM NaCl, SCP (pH 7.4) (red) and (b) 2 mg/mL (blue) and 1 mg/mL (black) LRAP(+P) in 2% acetic acid (pH 2.7) with the 2 mg/mL data normalized to the 1 mg/mL data to allow better comparison of the curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

R_g calculations for LRAP(+P) based on $P(r)$ analysis of SANS data with $D_{\max} = 70 \text{ \AA} \pm 5 \text{ \AA}$.

Condition	pH	[LRAP(+P)] mg/mL	R_g (nm)	Measured $I(0)$ (cm^{-1})	Calculated $I(0)$ (cm^{-1}) ^a	Apparent M_w (kDa)
2% (v/v) AA	2.7	2.0	2.06 ± 0.01	0.0114 ± 0.0004	0.0152	5.3 ± 0.2
2% (v/v) AA	2.7	1.0	2.05 ± 0.06	0.0067 ± 0.0003	0.0076	6.2 ± 0.2
SCP	7.4	1.0	2.06 ± 0.06	0.0067 ± 0.0002	0.0076	6.2 ± 0.2

^a $I(0)$ was calculated using Eq. (3) with $d = 1.34 \text{ g/cm}^3$ and assuming the LRAP monomer was measured in 100% D_2O solution. The LRAP sequence was taken from the Rosetta model PDB file. The monomer $M_w = 6.9 \text{ kDa}$, assuming 90% of exchangeable H are exchanged for D.

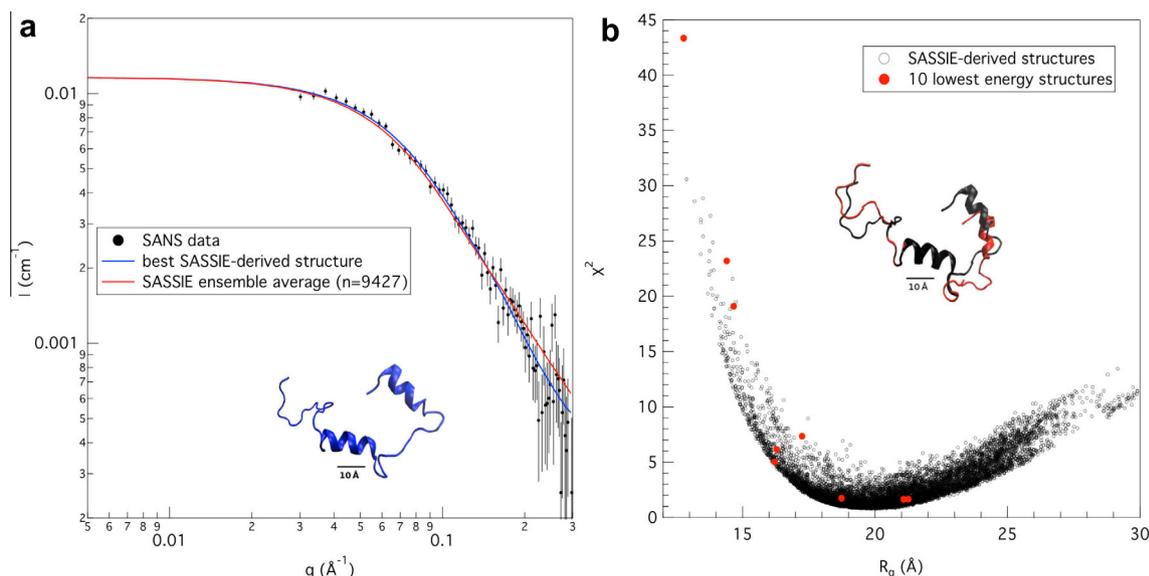


Fig. 6. (a) The SANS data from 2 mg/mL LRAP(+P) in 2% acetic acid (pH 2.7) along with the corresponding average (red) and best-fit (blue) model trace from the SASSIE calculations. The best-fit (blue) LRAP(+P) structure is shown under the traces. (b) A χ^2 versus R_g plot for the ensemble of 9427 SASSIE structures derived from the SANS data for 2 mg/mL LRAP(+P) in 2% acetic acid (pH 2.7). The corresponding points for the 10 lowest energy Rosetta model structures are shown in red. The best-fit (black) SASSIE structure and best-fit (red) lowest energy Rosetta structure are superimposed above the traces. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

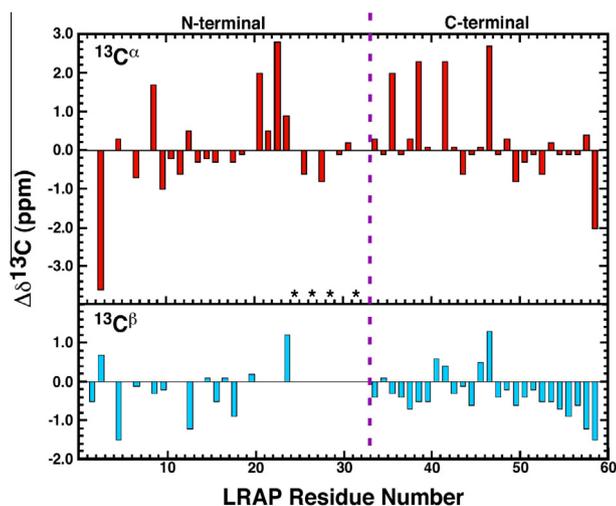


Fig. 7. Analysis of the observed $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift deviations from random coil values for rLRAP(-P) in 150 mM NaCl, pH 7.4 SCP solution where $\Delta\delta^{13}\text{C} = \delta_{\text{Random coil}} - \delta_{\text{Observed}}$. Red = $^{13}\text{C}^\alpha$, cyan = $^{13}\text{C}^\beta$. Asterisks indicate residues where it was not possible to obtain unambiguous $^{13}\text{C}^\beta$ assignments. The dashed purple line separated the residues present in the N- and C-terminal ends of full-length murine amelogenin. Random coil values obtained from `cns_solve_1.1`. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

isoelectric point where LRAP(+P) has a low charge and aggregates due to the loss of electrostatic repulsive forces. LRAP(+P), therefore, is a monomer at pH 3, aggregates at low charge in a pH range of

4.2–4.5, then forms monomers at pH values from 4.5 to 8.0, as shown in Fig. 4a. This behavior is in sharp contrast to full-length amelogenin which can form a range of quaternary structures as a function of pH. Amelogenin exists as monomers at pH 3, oligomers at pH 5.5, and nanospheres, self-assemblies of oligomers, at pH 6.8 and higher (Bromley et al., 2011).

The SV data reveals that LRAP(-P) has the same monomeric quaternary structure, size, and shape as LRAP(+P) at pH 7.4. Since the solution structures of LRAP(-P) and LRAP(+P) appear similar, further discussion will refer to LRAP in a general way. Although monomers are the dominant species in LRAP solutions at pH values higher than the isoelectric point, the solutions also contain a relatively low concentration of oligomeric species (0–14%). Our studies suggest that the oligomers are derived from the larger aggregates that are formed near the isoelectric point at $\sim\text{pH } 4.1$ as the solution pH is increased from the starting pH of ~ 3.0 . When the pH is increased above the isoelectric point, LRAP develops a negative charge which promotes dissolution of the aggregates to monomers and a small fraction of oligomers as small as dimers and trimers. When we change the way the solutions are made and dilute LRAP into basic buffer, avoiding the low charge pH regime, we find that the oligomer concentrations are greatly reduced.

Because the oligomers are present in relatively low concentrations, vary in concentration and size sample to sample, and are not present when the aggregation pH regime is avoided during sample preparation, we suggest that these species represent non-equilibrium structures. They are derived from aggregates formed near the isoelectric point of LRAP but the aggregates do not dissolve completely to the equilibrium monomeric species within

the time period of our measurements. We conclude, therefore, that LRAP does not form stable nanospheres under physiological conditions (pH 5.8–8.0). The lack of nanosphere formation for LRAP is consistent with a previous study on LRAP interactions using a yeast two-hybrid assay (Paine and Snead, 1997).

The differences in behavior between LRAP and amelogenin may lead to clues as to what protein–protein interactions are important in self-assembly. Table 1 shows that murine LRAP consists of the N-terminal residues M1–P33 and C-terminal residues P155–D180 of murine amelogenin but lacks the large central Y34–Q154 region. This suggests that the missing central region is necessary for nanosphere formation. This central region is very hydrophobic, rich in histidine (H, 13 residues), proline (P, 29 residues), and glutamine (Q, 23 residues). Fluorescence experiments of single tryptophan porcine amelogenin mutants showed that there are protein–protein interactions in the W25 and W45 region that increase as amelogenin forms oligomers at pH 5.5 and nanospheres at pH 8.0 (Bromley et al., 2011). The W45 region is located in the N-terminal side of the central domain that is missing in LRAP. The authors suggest that interactions near W25 and W45 promote oligomer formation at pH 5.5 and that further deprotonation of histidine residues in the central domain as the pH is raised leads to aggregation of the oligomers to form nanospheres. Hydrophobic interactions involving the central domain may also promote oligomer–oligomer binding.

Although the central hydrophobic region is necessary to promote oligomer and nanosphere formation in full-length amelogenin, the N and C-terminal domains have been proposed to have important roles in controlling the interactions (Buchko et al., 2008), size (Moradian-Oldak, 2000), polydispersity (Moradian-Oldak, 2000), and hierarchical structure (Fang et al., 2011) of the nanospheres. For example, removal of the N-terminal domain resulted in polydisperse aggregates ranging in size from 3 to 38 nm radius in contrast to the 20 nm radius monodisperse nanospheres of full-length amelogenin (Moradian-Oldak, 2000). Removal of the C-terminus resulted in larger aggregates (49 nm radius). Fang et al. (2011) proposed that the C-terminus may be important in the self-assembly of amelogenin into a double-ring barrel dodecamer as an intermediate to the formation of nanospheres. Amelogenin that was lacking the C-terminus (rM166) did not form the dodecamer oligomer intermediate.

4.2. LRAP tertiary structure (size and shape)

The SASSIE simulations show that LRAP has limited tertiary folding and is flexible enough to have a number of possible tertiary conformations. The hydrodynamic radius, R_h , from SV and radius of gyration, R_g , from SANS are both ~ 2.0 nm. This size is consistent with the smallest LRAP sizes obtained previously by DLS (Le Norcy et al., 2011). Both SV and SANS show that LRAP is extended with an asymmetric structure. For example, the SV ff/f_0 ratio of ~ 1.6 is typical of asymmetric proteins such as antibodies and indicates a significantly more extended or flexible structure than globular proteins with ff/f_0 ratios of 1.2–1.3 (Schuck, 2000). The SASSIE simulations from the SANS data shown in Fig. 6 also show that LRAP has an asymmetric structure. The size and shape of LRAP is the same at pH 2.7 and pH 7.4 according to the SANS studies. This suggests that the small increases in the ff/f_0 ratio from 1.59 at pH 8 to 1.64 at pH 3 may be due to slight increases in hydration and not due to any significant changes in shape. The relatively unchanged quaternary and tertiary structure of LRAP as a function of solution pH may be important to its function in an enamel environment of changing solution conditions.

4.3. LRAP secondary structure

Solution NMR studies of protein secondary structure show that rLRAP(–P) is predominantly unstructured, consistent with

previous CD studies of LRAP (Le et al., 2006; Chen et al., 2007). SSNMR studies have also shown evidence for random coil structures in the C-terminus and random coil to loose helices in the inner N-terminus of the protein when lyophilized from solution (Masica et al., 2011; Lu et al., 2013; Tarasevich et al., 2013). Rosetta simulations showed that the C-terminus (48–55) and inner N-terminus (10–20) had regions that were partially helical, sampling both helical and turn-like conformations (Tarasevich et al., 2013). The solution NMR studies presented here suggest that any loose or partially helical conformations in the terminal regions observed in the solid state or by simulation are likely transient in nature when LRAP is in solution. The Rosetta simulations (Masica et al., 2011) especially overestimate the degree of helical structure in the C and N-terminal regions compared to the experimental studies. The most structured part of LRAP is a stable helix localized in the L23–R31 region as evidenced by both SSNMR experiments and Rosetta simulations (Masica et al., 2011). This region, however, was experimentally inaccessible by the solution state NMR experiments described here.

Although only rLRAP(–P) was studied by solution NMR in this study, our CD studies (Fig. S7) show that LRAP(+P) is also predominantly unstructured. The local secondary structures of both LRAP(+P) and LRAP(–P) in the lyophilized state were also studied by SSNMR measurements of the ^{13}C – ^{15}N backbone distance (Masica et al., 2011; Lu et al., 2013). There were no significant differences in structure between LRAP(+P) and LRAP(–P) in the V19–L23 and K24–S28 regions and moderate differences in the G8–Y12 and K15–V19 regions. Interestingly, our CD studies show no evidence of secondary structure for LRAP(+P) in the presence of calcium even though our SV studies and previous SAXS studies (Le Norcy et al., 2012) find that LRAP has a more collapsed tertiary structure in the presence of calcium. This result suggests that calcium does not induce a random coil to α -helix or β -sheet transition for LRAP, consistent with previous studies (Le et al., 2006).

4.4. Physiological significance

The general LRAP secondary, tertiary, and quaternary structures are similar to those of the SIBLING family of biomineralization proteins, known to be involved in the formation of bone and dentin (George and Veis, 2008). These glycoproteins include bone sialoprotein (BSP), dental phosphophoryn (DPP), dentin matrix protein 1 (DMP1), and dentin sialoprotein (DSP). The SIBLING proteins exist as monomers, have flexible backbones, and have limited global secondary or tertiary structure (Fisher et al., 2001). There is evidence that the SIBLING proteins can have dual roles in bone and dentin formation – an extracellular matrix function in guiding crystal nucleation and growth, and a cell signaling function (George and Veis, 2008). For example, BSP has been suggested to be the main nucleator of HAP in bone (Ganss et al., 1999) as well as a promoter of osteoblast differentiation (Gordon et al., 2007). Like the SIBLING proteins, the monomeric, flexible, and relatively unstructured nature of LRAP may be important in its function as a cell signaling molecule, promoting binding interactions with cell surface receptors. A flexible, monomeric protein conformation may also be important to an extracellular crystal growth function, allowing adaptation of the protein to the growing crystal lattice. The SV, SANS, and NMR techniques developed here to elucidate the structure of LRAP can also be applied to studying other amelogenin proteins including other splice variants and truncated forms of LRAP (Li et al., 1999; Nagano et al., 2009).

5. Summary

Our studies have shown that analytical techniques such as SV, SANS, and NMR can provide important insights into the structure

and function of biomineralization proteins. We found that LRAP does not form nanospheres like the higher molecular weight amelogenins and exists as a monomer over a range of solution conditions. This result supports the hypothesis that the central hydrophobic domain present in full-length amelogenin is critical to self-assembly. The LRAP monomer has an asymmetric shape and an equivalent spherical radius of ~ 2 nm with no significant changes as a function of pH. The monomer is more collapsed at higher salt concentrations and in the presence of calcium. The protein is extended and flexible with limited tertiary folding and a limited degree of local secondary structure. LRAP's monomeric, largely unstructured conformation may be important to its biological function as a cell signaling or extracellular matrix protein during enamel formation as well as to a therapeutic function for the regeneration of periodontal tissue.

Acknowledgments

This work was supported by NIH-NIDCR Grant DE-015347 (PNNL) and was performed in part at Pacific Northwest National Laboratory, operated by Battelle for the US-DOE. A portion of the research was performed in the EMSL, a national scientific user facility sponsored by the DOE-OBER at PNNL. The SANS studies benefitted from CCP-SAS software developed by Joseph Curtis through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jsb.2014.10.007>.

References

- Amin, H.D., Olsen, I., Knowles, J.C., Donos, N., 2012. Differential effect of amelogenin peptides on osteogenic differentiation in vitro: identification of possible new drugs for bone repair and regeneration. *Tissue Eng. Part A* 18, 1193–1202.
- Boabaid, F., Gibson, C.W., Kuehl, M.A., Berry, J.E., Snead, M.L., Nociti, F.H.J., Katchburian, E., Somerman, M.J., 2004. Leucine-rich amelogenin peptide: a candidate signaling molecule during cementogenesis. *J. Periodontol.* 75, 1126–1136.
- Bromley, K.M., Kiss, A.S., Lokappa, S.B., Lakshminarayanan, R., Fan, D., Ndao, M., Evans, J.S., Moradian-Oldak, J., 2011. Dissecting amelogenin protein nanospheres: characterization of metastable oligomers. *J. Biol. Chem.* 286, 34643–34653.
- Buchko, G.W., Tarasevich, B.J., Bekhazi, J., Snead, M.L., Shaw, W.J., 2008. A solution NMR investigation into the early events of amelogenin nanosphere self-assembly initiated with sodium chloride or calcium chloride. *Biochemistry* 47, 13215–13222.
- Buchko, G.W., Tarasevich, B.J., Roberts, J., Snead, M.L., Shaw, W.J., 2010. A solution NMR investigation into the murine amelogenin splice-variant LRAP (Leucine-Rich Amelogenin Protein). *Biochim. Biophys. Acta* 1804, 1768–1774.
- Chen, X., Wang, Q., Shen, J., Pan, H., Wu, T., 2007. Adsorption of leucine-rich amelogenin protein on hydroxyapatite (001) surface through $-\text{COO}^-$ claws. *J. Phys. Chem. C* 111, 1284–1290.
- Curtis, Raghunandan, Krueger, Nanda, 2012. SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comput. Phys. Commun.* 183, 8–8.
- Delak, K., Harcup, C., Lakshminarayanan, R., Sun, Z., Fan, Y., Moradian-Oldak, J., Evans, J.S., 2009. The tooth enamel protein, porcine amelogenin, is an intrinsically disordered protein with an extended molecular configuration in the monomeric form. *Biochemistry* 48, 2272–2281.
- Fang, P.-A., Conway, J.F., Margolis, H.C., Simmer, J.P., Beniash, E., 2011. Hierarchical self-assembly of amelogenin and the regulation of biomineralization at the nanoscale. *Proc. Natl. Acad. Sci. USA* 108, 14097–14102.
- Filipe, V., Hawe, A., Jiskoot, W., 2010. Critical evaluation of nanoparticle tracking analysis (NTA) by nanosight for the measurement of nanoparticles and protein aggregates. *Pharm. Res.* 27, 796–810.
- Fincham, A.G., Moradian-Oldak, J., Diekwisch, T.G., Lyaruu, D.M., Wright, J.T., Bringas, P., Slavkin, H.C., 1995. Evidence for amelogenin “nanospheres” as functional components of secretory-stage enamel matrix. *J. Struct. Biol.* 115, 50–59.
- Fincham, A.G., Moradian-Oldak, J., Simmer, J.P., 1999. The structural biology of the developing dental enamel matrix. *J. Struct. Biol.* 126, 30–30.
- Fisher, L.W., Torchia, D.A., Fohr, B., Young, M.F., Fedarko, N.S., 2001. Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochem. Biophys. Res. Commun.* 280, 460–465.
- Ganss, B., Kim, R.H., Sodek, J., 1999. Bone sialoprotein. *Crit. Rev. Oral Biol. Med.* 10, 79–98.
- George, A., Veis, A., 2008. Phosphorylated proteins and control over apatite nucleation, crystal growth, and inhibition. *Chem. Rev.* 108, 4670–4693.
- Gibson, C.W., 2011. The amelogenin proteins and enamel development in humans and mice. *J. Oral Biosci.* 53, 248–256.
- Gibson, C.W., Golub, E., Ding, W., Shimokawa, H., Young, M., Termine, J., Rosenbloom, J., 1991. Identification of the leucine-rich amelogenin peptide (LRAP) as the translation product of an alternatively spliced transcript. *Biochem. Biophys. Res. Commun.* 174, 1306–1312.
- Gibson, C.W., Kucich, U., Collier, P., Shen, G., Decker, S., Bashir, M., Rosenbloom, J., 1995. Analysis of amelogenin proteins using monospecific antibodies to defined sequences. *Connect. Tissue Res.* 32, 109–114.
- 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., Kulkarni, A.B., 2001. Amelogenin-deficient mice display an amelogenesis imperfecta phenotype. *J. Biol. Chem.* 276, 31871–31875.
- Gibson, C.W., Li, Y., Daly, B., Suggs, C., Yuan, Z.-A., Fong, H., Simmons, D., Aragon, M., Kulkarni, A.B., Wright, J.T., 2009. The leucine-rich amelogenin peptide alters the amelogenin null enamel phenotype. *Cells Tissues Organs* 189, 169–174.
- Gibson, C.W., Li, Y., Suggs, C., Kuehl, M.A., Pugach, M.K., Kulkarni, A.B., Wright, J.T., 2011. Rescue of the murine amelogenin null phenotype with two amelogenin transgenes. *Eur. J. Oral Sci.* 119 (Suppl. 1), 70–74.
- Glinka, C.J., Barker, J.G., Hammouda, B., Krueger, S., Moyer, J.J., Orts, W.J., 1998. The 30 m small-angle neutron scattering instruments at the national institute of standards and technology. *J. Appl. Crystallogr.* 31, 430–445.
- Gordon, J.A.R., Tye, C.E., Sampaio, A.V., Underhill, T.M., Hunter, G.K., Goldberg, H.A., 2007. Bone sialoprotein expression enhances osteoblast differentiation and matrix mineralization in vitro. *Bone* 41, 462–473.
- Habelitz, S., DenBesten, P.K., Marshall, S.J., Marshall, G.W., Li, W., 2006. Self-assembly and effect on crystal growth of the leucine-rich amelogenin peptide. *Eur. J. Oral Sci.* 114, 315–319.
- Hammarmstrom, L., 1997. The role of enamel matrix proteins in the development of cementum and periodontal tissues. *Ciba Found. Symp.* 205, 246–255 (discussion 255–260).
- Heidorn, D.B.D., Trehwella, J.J., 1988. Comparison of the crystal and solution structures of calmodulin and troponin C. *Biochemistry* 27, 909–915.
- Heijl, L.L., Heden, G.G., Svärdröm, G.G., Ostgren, A.A., 1997. Enamel matrix derivative (EMDOGAIN) in the treatment of intrabony periodontal defects. *J. Clin. Periodontol.* 24, 705–714.
- Iijima, M., Moradian-Oldak, J., 2004. Control of octacalcium phosphate and apatite crystal growth by amelogenin matrices. *J. Mater. Chem.* 14, 2189.
- Kline, S.R., 2006. Reduction and analysis of SANS and USANS data using IGOR Pro. *J. Appl. Crystallogr.* 39, 895–900.
- Krueger, S.S., Gorshkova, I.I., Brown, J.J., Hoskins, J.J., McKenney, K.H.K., Schwarz, F.P.F., 1998. Determination of the conformations of cAMP receptor protein and its T127L, S128A mutant with and without cAMP from small angle neutron scattering measurements. *J. Biol. Chem.* 273, 20001–20006.
- Laue, T.M., Shah, B.D., Ridgeway, T.M., Pelletier, S.L., 1992. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding, S.E., Rowe, A.J., Horton, J.C. (Eds.), *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Society of Chemistry, Cambridge, pp. 90–125.
- Le, T.Q., Gochin, M., Featherstone, J.D.B., Li, W., DenBesten, P.K., 2006. Comparative calcium binding of leucine-rich amelogenin peptide and full-length amelogenin. *Eur. J. Oral Sci.* 114 (Suppl. 1), 320–326 (discussion 327–9, 382).
- Le, T.Q., Zhang, Y., Li, W., DenBesten, P.K., 2007. The effect of LRAP on enamel organ epithelial cell differentiation. *J. Dent. Res.* 86, 1095–1099.
- Le Norcy, E., Kwak, S.Y., Wiedemann-Bidlack, F.B., Beniash, E., Yamakoshi, Y., Simmer, J.P., Margolis, H.C., 2011. Leucine-rich amelogenin peptides regulate mineralization in vitro. *J. Dent. Res.* 90, 1091–1097.
- Le Norcy, E., Kwak, S.-Y., Allaire, M., Fratzl, P., Yamakoshi, Y., Simmer, J.P., Margolis, H.C., 2012. Effect of phosphorylation on the interaction of calcium with leucine-rich amelogenin peptide. *Eur. J. Oral Sci.* 119 (Suppl. 1), 97–102.
- Lebowitz, J.J., Lewis, M.S.M., Schuck, P.P., 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* 11, 2067–2079.
- Li, W., Machule, D., Gao, C., DenBesten, P.K., 1999. Activation of recombinant bovine matrix metalloproteinase-20 and its hydrolysis of two amelogenin oligopeptides. *Eur. J. Oral Sci.* 107, 352–359.
- Lu, J.-X., Xu, Y.S., Shaw, W.J., 2013. Phosphorylation and ionic strength alter the LRAP-HAP interface in the N-terminus. *Biochemistry* 52, 2196–2205.
- MacKerell, A.D., Bashford, D., Bellott, Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616.
- Masica, D.L., Gray, J.J., Shaw, W.J., 2011. Partial high-resolution structure of phosphorylated and non-phosphorylated leucine-rich amelogenin protein adsorbed to hydroxyapatite. *J. Phys. Chem. C* 115, 13775–13785.

- Moradian-Oldak, J., 2000. Self-assembly properties of recombinant engineered amelogenin proteins analyzed by dynamic light scattering and atomic force microscopy. *J. Struct. Biol.* 131, 27–37.
- Moradian-Oldak, J., 2001. Amelogenins: assembly, processing and control of crystal morphology. *Matrix Biol.* 20, 293–305.
- Moradian-Oldak, J., Simmer, J.P., Lau, E.C., Sarte, P.E., Slavkin, H.C., Fincham, A.G., 1994. Detection of monodisperse aggregates of a recombinant amelogenin by dynamic light scattering. *Biopolymers* 34, 1339–1347.
- Moradian-Oldak, J., Leung, W., Fincham, A.G., 1998. Temperature and pH-dependent supramolecular self-assembly of amelogenin molecules: a dynamic light-scattering analysis. *J. Struct. Biol.* 122, 320–327.
- Moradian-Oldak, J., Tan, J., Fincham, A.G., 1998a. Interaction of amelogenin with hydroxyapatite crystals: an adherence effect through amelogenin molecular self-association. *Biopolymers* 46, 225–238.
- Moradian-Oldak, J., Tan, J., Fincham, A.G., 1998b. Interaction of amelogenin with hydroxyapatite crystals: an adherence effect through amelogenin molecular self-association. *Biopolymers* 46, 225–238.
- Nagano, T., Kakegawa, A., Yamakoshi, Y., Tsuchiya, S., Hu, J.C.C., Gomi, K., Arai, T., Bartlett, J.D., Simmer, J.P., 2009. Mmp-20 and Kik4 cleavage site preferences for amelogenin sequences. *J. Dent. Res.* 88, 823–828.
- Paine, M.L., Snead, M.L., 1997. Protein interactions during assembly of the enamel organic extracellular matrix. *J. Bone Miner. Res.* 12, 221–227.
- Phillips, J.C.J., Braun, R.R., Wang, W.W., Gumbart, J.J., Tajkhorshid, E.E., Villa, E.E., Chipot, C.C., Skeel, R.D.R., Kalé, L.L., Schulten, K.K., 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802.
- Philo, J.S., 2006. Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal. Biochem.* 354, 238–246.
- Ravindranath, R.M.H., Devarajan, A., Bringas Jr., P., 2007. Enamel formation in vitro in mouse molar explants exposed to amelogenin polypeptides. *Arch. Oral Biol.* 52, 1161–1171.
- Robinson, C., Kirkham, J., Brookes, S.J., Bonass, W.A., Shore, R.C., 1995. The chemistry of enamel development. *Int. J. Dev. Biol.* 39, 145–152.
- Sasaki, S.S., Takagi, T.T., Suzuki, M.M., 1991. Cyclical changes in pH in bovine developing enamel as sequential bands. *Arch. Oral Biol.* 36, 227–231.
- Schuck, P.P., 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* 78, 14–14.
- Sculean, A.A., Kiss, A.A., Miliauskaitė, A.A., Schwarz, F.F., Arweiler, N.B.N., Hannig, M.M., 2008. Ten-year results following treatment of intra-bony defects with enamel matrix proteins and guided tissue regeneration. *J. Clin. Periodontol.* 35, 817–824.
- Semenyuk, A.V., Svergun, D.I., 1991. GNOM – a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* 24, 537–540.
- Shaw, W.J., Campbell, A.A., Paine, M.L., Snead, M.L., 2004. The COOH terminus of the amelogenin, LRAP, is oriented next to the hydroxyapatite surface. *J. Biol. Chem.* 279, 40263–40266.
- Stahl, J., Nakano, Y., Kim, S.-O., Gibson, C.W., Le, T., Denbesten, P., 2013. Leucine rich amelogenin peptide alters ameloblast differentiation in vivo. *Matrix Biol.* 32, 432–442.
- Tan, J., Leung, W., Moradian-Oldak, J., Zeichner-David, M., Fincham, A.G., 1998. Quantitative analysis of amelogenin solubility. *J. Dent. Res.* 77, 1388–1396.
- Tarasevich, B.J., Howard, C.J., Larson, J.L., Snead, M.L., Simmer, J.P., Paine, M., Shaw, W.J., 2007. The nucleation and growth of calcium phosphate by amelogenin. *J. Cryst. Growth* 304, 407–415.
- Tarasevich, B.J., Lea, S., Shaw, W.J., 2010. The leucine rich amelogenin protein (LRAP) adsorbs as monomers or dimers onto surfaces. *J. Struct. Biol.* 169, 266–276.
- Tarasevich, B.J., Perez-Salas, U., Masica, D.L., Philo, J., Kienzle, P., Krueger, S., Majkrzak, C.F., Gray, J.L., Shaw, W.J., 2013. Neutron reflectometry studies of the adsorbed structure of the amelogenin, LRAP. *J. Phys. Chem. B* 117, 3098–3109.
- Termine, J.D., Belcourt, A.B., Christner, P.J., Conn, K.M., Nylen, M.U., 1980. Properties of dissociatively extracted fetal tooth matrix proteins. I. Principal molecular species in developing bovine enamel. *J. Biol. Chem.* 255, 9760–9768.
- Veis, A., Tompkins, K., Alvares, K., Wei, K., Wang, L., Wang, X.S., Brownell, A.G., Jengh, S.M., Healy, K.E., 2000. Specific amelogenin gene splice products have signaling effects on cells in culture and in implants in vivo. *J. Biol. Chem.* 275, 41263–41272.
- Warotayanont, R., Zhu, D., Snead, M.L., Zhou, Y., 2008. Leucine-rich amelogenin peptide induces osteogenesis in mouse embryonic stem cells. *Biochem. Biophys. Res. Commun.* 367, 1–6.
- Warotayanont, R., Frenkel, B., Snead, M.L., Zhou, Y., 2009. Leucine-rich amelogenin peptide induces osteogenesis by activation of the Wnt pathway. *Biochem. Biophys. Res. Commun.* 387, 558–563.
- Wen, X., Cawthorn, W.P., MacDougald, O.A., Stupp, S.I., Snead, M.L., Zhou, Y., 2011. The influence of leucine-rich amelogenin peptide on MSC fate by inducing Wnt10b expression. *Biomaterials* 32, 6478–6486.
- Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., Sykes, B.D., 1995. ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6, 135–140.