bio-techne®

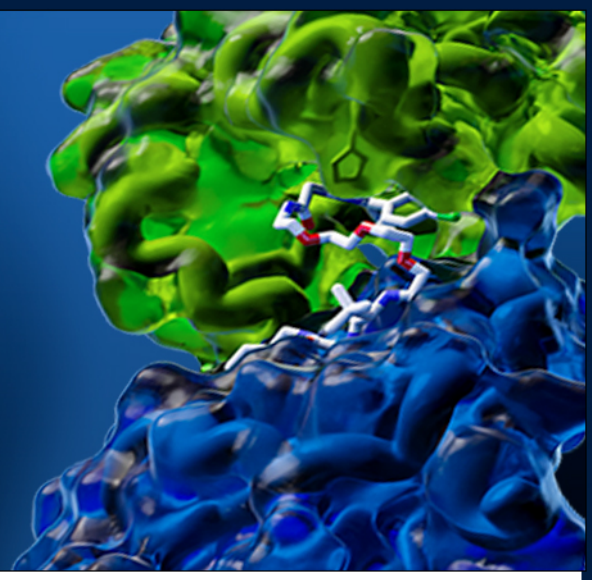
Analytical Exchange Webinar

Accelerating Protein Analysis Throughput at Boehringer Ingelheim

October 3, 2023

3pm BST | 4pm CEST | 10am EDT | 7am PDT

Register



Advancing Drug Discovery with Targeted Protein Degradation

Unlike traditional drug discovery, which focuses on inhibiting or activating proteins, targeted protein degradation (TPD) offers a more precise and efficient way to alter cellular pathways. However, TPD faces several challenges that need to be overcome to reach its full potential, including reducing the bottleneck of quantitative protein analysis.

In this upcoming webinar, the team at Boehringer Ingelheim present their recent real- world TPD project examples and in-depth analysis, covering insights on device and consumable management and optimized assay setups to increase protein analysis throughput.

Speakers

Andrea Stingu
Lab Scientist
Boehringer Ingelheim

Johannes Wachter
Lab Scientist
Boehringer Ingelheim

Teressa Puchner
Lab Scientist
Boehringer Ingelheim

Register Now

Bio-Techne
614 McKinley Place NE, Minneapolis, MN USA 55413

If you no longer wish to receive messages from Bio-Techne you can unsubscribe at any time. You can update your communication preferences in our preference center.
Copyright © 2023 Bio-Techne. All rights reserved.

Check for updates

Received: 21 October 2022 | Revised: 21 December 2022 | Accepted: 29 December 2022

DOI: 10.1002/pro.4560

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

High-yield recombinant bacterial expression of ^{13}C-, ^{15}N-labeled, serine-16 phosphorylated, murine amelogenin using a modified third generation genetic code expansion protocol

Garry W. Buchko[1,2] | **Mowei Zhou**[1] | **Cat Hoang Vesely**[3] | **Jinhui Tao**[4] | **Wendy J. Shaw**[4] | **Ryan A. Mehl**[3] | **Richard B. Cooley**[3]

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, USA

[2]School of Molecular Biosciences, Washington State University, Pullman, Washington, USA

[3]Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon, USA

[4]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, USA

Correspondence
Garry W. Buchko, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA.
Email: garry.buchko@pnnl.gov

Richard B. Cooley, Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA.
Email: rick.cooley@oregonstate.edu

Funding information
National Institute of Dental and Craniofacial Research, Grant/Award Number: DE-015347

Review Editor: John Kuriyan

Abstract

Amelogenin constitutes ~90% of the enamel matrix in the secretory stage of amelogenesis, a still poorly understood process that results in the formation of the hardest and most mineralized tissue in vertebrates—enamel. Most biophysical research with amelogenin uses recombinant protein expressed in *Escherichia coli*. In addition to providing copious amounts of protein, recombinant expression allows ^{13}C- and ^{15}N-labeling for detailed structural studies using NMR spectroscopy. However, native amelogenin is phosphorylated at one position, Ser-16 in murine amelogenin, and there is mounting evidence that Ser-16 phosphorylation is important. Using a modified genetic code expansion protocol we have expressed and purified uniformly ^{13}C-, ^{15}N-labeled murine amelogenin (pS16M179) with ~95% of the protein being correctly phosphorylated. Homogeneous phosphorylation was achieved using commercially available, enriched, ^{13}C-, ^{15}N-labeled media, and protein expression was induced with isopropyl β-D-1-thiogalactopyranoside at 310 K. Phosphoserine incorporation was verified from one-dimensional ^{31}P NMR spectra, comparison of ^{1}H-^{15}N HSQC spectra, Phos-tag SDS PAGE, and mass spectrometry. Phosphorus-31 NMR spectra for pS16M179 under conditions known to trigger amelogenin self-assembly into nanospheres confirm nanosphere models with buried N-termini. Lambda phosphatase treatment of these nanospheres results in the dephosphorylation of pS16M179, confirming that smaller oligomers and monomers with exposed N-termini are in equilibrium with nanospheres. Such ^{13}C-, ^{15}N-labeling of amelogenin with accurately encoded phosphoserine

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 Battelle Memorial Institute and The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

incorporation will accelerate biomineralization research to understand amelogenesis and stimulate the expanded use of genetic code expansion protocols to introduce phosphorylated amino acids into proteins.

KEYWORDS

^{31}P-NMR, amelogenesis, amelogenin, biomineralization, isotope labeling, nanospheres, phosphoproteins, posttranslational modifications, recombinant protein expression

1 | INTRODUCTION

Protein phosphorylation is the most common posttranslational modification cells use to regulate protein function (Pawson & Scott, 2005; Ubersax & Ferrell, 2007). Included among the basic cellular processes regulated by protein phosphorylation is amelogenesis (Shin et al., 2020), an activity that results in the formation of the hardest, toughest, and most mineralized tissue in vertebrates––enamel (Simmer & Fincham, 1995; Ten Cate, 1994). Enamel is composed of narrow crystals of carbonated hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$) and its extraordinary mechanical properties arise from its weaving into an unique, closely packed, lattice architecture (Daculsi et al., 1984; Hunter, 1996; Margolis et al., 2006). Amelogenesis occurs in specialized epithelial cells called ameloblasts and is regulated by the proteins amelogenin, ameloblastin, enamelin, and amelotin. Of these proteins, the dominant protein (>90%) at the secretory stage of amelogenesis is amelogenin (Margolis et al., 2006; Simmer & Fincham, 1995). Genetically modified and knock-out mouse models show amelogenin is a key component for proper enamel formation (Gibson et al., 2001; Hu et al., 2016). This importance is corroborated by human genetic studies showing mutations to the amelogenin gene AMELX is a leading cause of *amelogenesis imperfecta* (Ravassipour et al., 2000; Witkop Jr. et al., 1973), a heterogenous group of hereditary conditions that affect the quantity and quality of enamel (Smith et al., 2017). The AMELX gene product contains only one posttranslational modification, the phosphorylation of a single serine residue near the N-terminal which in human, mice, and pig amelogenin is Ser-16 (pSer-16).

Analysis of the primary amino acid sequence of amelogenin from humans and other vertebrate tetrapods shows that the sequence is highly homologous across species (Toyosawa et al., 1998) and can be divided into three domains: an N-terminal, hydrophilic tyrosine-rich region (a.k.a. TRAP), a hydrophobic central region heavily populated with histidine, glutamine, and proline residues, and a hydrophilic mineral binding C-terminal region (Margolis et al., 2006; Shaw et al., 2020). The N- and C-terminal domains are highly conserved across species with small species-dependent variations in the length of amelogenin due largely to differences in the length of the central domain. Structural studies suggest amelogenin is an intrinsically disordered protein especially in acidic aqueous solution under dilute conditions where it is primarily monomeric (Buchko, Tarasevich, et al., 2008; Delak et al., 2009). Perhaps due to the histidine residues in the central domain (Bromley et al., 2011), as the solution pH is increased amelogenin self-assembles into different quaternary structures in a dynamic, step-wise fashion (Bromley et al., 2011; Lakshminarayanan et al., 2007; Moradian-Oldak et al., 1994; Moradian-Oldak et al., 1998). Starting at low pH (<~3.5) ameloginin is primarily monomeric (Buchko, Tarasevich, et al., 2008; Delak et al., 2009). As physiological pHs are approached (pH ~6.6) amelogenin forms oligomers of increasing size up to a maximum average size of about eight monomers (Du et al., 2005; Fang et al., 2011). At pH values above ~7.2, nanospheres composed of 20–100 monomers form (Bromley et al., 2011; Du et al., 2005) and, under the right conditions, nanosphere chains (Wiedemann-Bidlack et al., 2007) and nanoribbons (Carneiro et al., 2016; He et al., 2011). Nanospheres have been observed both *in vitro* (Fincham et al., 1994) and *in vivo* (Fincham et al., 1995), and appear to be essential for proper enamel formation (Paine et al., 2000). In addition to pH, the equilibrium between various quaternary structures is influenced by other factors including the amelogenin concentration, the presence of amelogenin degradation products, and the properties of the solution such as ionic strength, solutes, and temperature (Engelberth et al., 2018; Moradian-Oldak et al., 1998; Shaw et al., 2020).

While there is little doubt that amelogenin plays a critical role in amelogenesis, how this ~20 kDa, intrinsically disordered protein controls hydroxyapatite (HAP) crystal growth at the molecular level is still not well understood. A contributing factor toward this knowledge gap may be because a significant amount of the biophysical research on amelogenin has been conducted using recombinant protein expressed in *Escherichia coli*. While recombinant methods have been successful in providing copious amounts of protein for various biophysical studies and ^{13}C- and ^{15}N-labeled protein for detailed structural studies using NMR spectroscopy (Buchko et al., 2022; Buchko, Bekhazi, et al., 2008; Buchko, Tarasevich, et al., 2008; Delak et al., 2009),

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

recombinant amelogenin expressed in *E. coli* lacks a phosphoserine. A few scattered experiments using native phosphorylated amelogenin (Fang et al., 2013; Wiedemann-Bidlack et al., 2011), peptides of native amelogenin digestion products (Kwak et al., 2009), or the amelogenin splice-variant LRAP chemically synthesized with pSer-16 (Lu et al., 2013; Yamazaki et al., 2017) have hinted that amelogenin phosphorylation may affect the physical properties of the protein and play a role in regulating enamel formation. Recently, Margolis et al. provided some of the strongest *in vivo* evidence to date for the importance of amelogenin phosphorylation using a hemizygous knock-in mouse model with a Ser-16 to Ala substitution in the AMELX gene (Shin et al., 2020). The enamel of such genetically engineered mice was defective, lacking enamel rods and exhibiting hypoplasia and numerous surface defects. To accelerate the biomineralization community's efforts to more fully explore the biophysical effects of amelogenin phosphorylation, we describe the application of the latest generation of genetic code expansion technology (Zhu et al., 2019) for obtaining milligram quantities of ^{13}C- and ^{15}N-labeled, Ser-16 phosphorylated, murine amelogenin (pS16M179) using recombinant methods in *E. coli*. The advantage of illustrating this method with ^{13}C- and ^{15}N-labeled protein is because the NMR chemical shifts for non-phosphorylated recombinant murine amelogenin (M179, a 179-residue protein after N-terminal methionine removal) are known (Buchko, Bekhazi, et al., 2008; Buchko, Tarasevich, et al., 2008), and consequently, ^{13}C- and ^{15}N-labelling allows us to readily verify the presence of Ser-16 phosphorylation, the extent of Ser-16 phosphorylation, and the stability of Ser-16 phosphorylation. Note that genetic code expansion technologies that encode for phosphoserine use engineered *E. coli* expression hosts that lack the *serB* gene. *SerB* encodes for a serine phosphatase that converts free phosphoserine amino acids into serine as part of the serine biosynthesis pathway. By deleting *serB*, free pSer builds up inside the cells and feeds the engineered pSer genetic code expansion system. Because these *serB*$^{-}$ *E. coli* expression hosts are serine auxotrophs, classical ^{13}C- and ^{15}N-labeling protocols using minimal media are ineffective (Zhu et al., 2019). Instead, it is necessary to use enriched, ^{13}C- and ^{15}N-labeled media, such as Celtone or BioExpress (Vesely et al., 2022).

2 | RESULTS AND DISCUSSION

2.1 | Expression of phosphorylated pS16M179

To genetically encode phosphoserine at amber stop codons in *E. coli* during translation, we employed a pSer genetic code expansion machinery system previously developed by Chin and colleagues (Rogerson et al., 2015). These expression systems rely on the use of expression hosts that lack the *serB* gene as a means to build up free phosphoserine amino acid inside the cell, which serves to feed the genetic code expansion machinery. Because *serB* is critical for serine biosynthesis, these *E. coli* expression hosts are serine auxotrophs and cannot grow in traditional, isotopically enriched minimal media. However, we observed that normal growth of these serine auxotrophs could be restored by supplementing minimal media with Celtone, an algal extract which can be commercially purchased enriched with various isotopes, including ^{15}N and ^{13}C (Vesely et al., 2022). Using such media we recently optimized a high-density culturing method using BL21(DE3) Δ*serB* as an expression host, which enabled for the first time, high yields of homogeneously phosphorylated, isotopically labeled proteins. However, our attempts to express phosphorylated amelogenin in high yields with this strain and these methods were not successful (data not shown).

We hypothesized that using a strain of *E. coli* called B95(DE3) Δ*A* Δ*fabR* Δ*serB*, which lacks Release Factor-1 (RF1, the protein responsible for terminating translation at amber codons) could improve incorporation by removing competition between the amber suppressing, pSer-amino acylated tRNA and RF1. The absence of RF1, however, opens the door for endogenous tRNAs (e.g Gln-tRNA) to suppress UAG codons via wobble-base pairing (Beyer et al., 2020; Zhu et al., 2019), a phenomenon referred to as near-cognate suppression. Therefore, checking the accuracy of encoding when using RF1 deficient expression hosts is particularly important to ensure the GCE machinery can effectively outcompete endogenous near-cognate suppressor tRNAs. Indeed, we previously demonstrated accurate encoding of pSer when using rich auto-induction media with the B95(DE3) Δ*A* Δ*fabR* Δ*serB* expression host (Zhu et al., 2019).

We therefore tested whether this cell line, B95(DE3) Δ*A* Δ*fabR* Δ*serB*, could be used to produce isotopically enriched, phosphorylated proteins using our high-density culturing methods previously optimized with BL21(DE3) Δ*serB*. Similar or improved yields of the model protein, super-folder green fluorescent protein with an amber (TAG) stop codon for phosphoserine at residue position 150 (sfGFP-150TAG) (Zhu et al., 2019) were obtaining using Phos-tag gel electrophoresis to assess the proportion of purified protein that was phosphorylated (Figure S1). Phos-tag gels contain a di-nuclear metal complex (1,3-bis[bis-(pyridine-2-ylmethyl)amino]propan-2-olato dizinc(II)) copolymerized in an acrylamide matrix with affinity for phosphate groups that retards phosphoprotein migration in the gel during electrophoresis

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Kinoshita et al., 2006). The method is sensitive enough to separate a protein by the number of sites phosphorylated (Kinoshita et al., 2006; Zhu et al., 2019). Unfortunately, an unacceptable proportion of the purified protein (up to 90%) was not phosphorylated using these high-density expression methods due to near-cognate suppression. Through rounds of further expression optimization, we found that reverting back to a low-density culturing method and expressions at elevated temperatures (298–310 K) notably improved the faithful encoding of pSer into sfGFP-150TAG (up to 90%) using the B95 (DE3) Δ*A* Δ*fabR* Δ*serB* expression host (Figure S1).

Guided by these improvements using sfGFP as a model protein, we asked whether the B95(DE3) Δ*A* Δ*fabR* Δ*serB* strain could be used to express phosphorylated amelogenin and tested a variety of culturing and expression conditions. Figure 1 shows a Phos-tag gel for purified pS16M179 obtained from ^{13}C-, ^{15}N-labeled Celtone complete and other media, and Table 1 summarizes the approximate yields of amelogenin using these media. A number of general conclusions can be made from an inspection of the gel. First, un-phosphorylated M179 clearly migrates faster than pS16M179, confirming the phosphorylation of the protein. Second, relative to M179, only one major slower migrating band is detected indicating that the phosphorylation is confined to one site. Third, based on a visual inspection of the gel, the level of pSer incorporation is high regardless of the media used during protein expression. Fourth, IPTG induction in ZY-media results in moderately lower levels of misincorporation for pSer than autoinduction in ZY-media, however, the yields of amelogenin are about 3× greater using autoinduction. Fifth, the level of misincorporation was greatest in the deuterated Celtone complete media compared to the media using 100% H_2O and there was a greater number of faster migrating impurities in the sample perhaps due to the slower growth rate of *E. coli* grown in D_2O (Opitz et al., 2019). For murine amelogenin at least, rapid growth appeared essential for the expression of pS16M179 as no protein expression was observed in ZY-media at 293 K (Table 1).

2.2 | NMR verification of pSer-16 incorporation into pS16M179

Additional indication that the genetic code expansion method for incorporating phosphoserine into M179 worked was the one-dimensional ^{31}P NMR spectrum for pS16M179 shown in Figure 2 (top). At pH 2.8 a single resonance at 0.7 ppm was observed which is near the chemical shift observed for the phosphoryl group in pSer at 0.5 ppm (bottom). As shown in Table 2, these ^{31}P chemical shifts are also near the chemical shift for the phosphorylated serine observed in a number of short, random coil peptides at pH 3 (Bienkiewicz & Lumb, 1999; Hoffmann et al., 1994), suggesting this region of pS16M179 is also disordered. This is in agreement with earlier studies showing non-phosphorylated murine and porcine amelogenin are intrinsically disordered at pH 2.8 (Buchko, Tarasevich, et al., 2008; Delak et al., 2009). Note that there were no noticeable changes to one-dimensional ^{31}P spectra of pS16M179 and pSer at pH 2.8 upon incubation of the samples in the NMR tubes at 293 K for over a month, indicating that serine phosphorylation was stable under these conditions over this

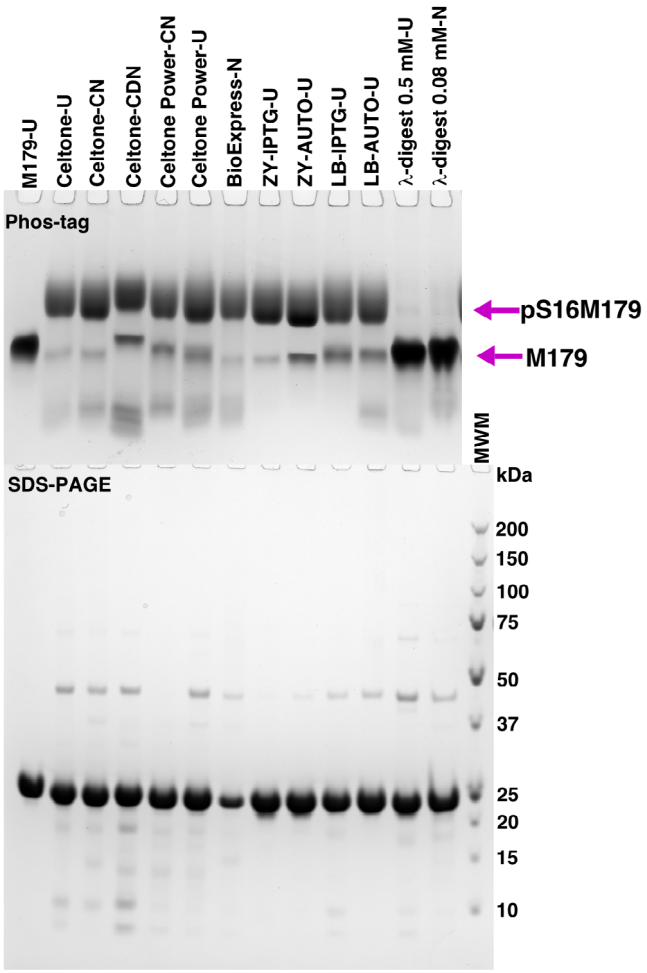


FIGURE 1 Coomassie-stained Phos-tag (top) and SDS-PAGE (bottom) gels for M179 and pS16M179 expressed in B21(DE3) and B95(DE3) *E. coli*, respectively, using a suite of media. The last two lanes are pS16M179 treated with lambda phosphorylase (λ) at two different concentrations (0.5 and 0.08 mM). Approximately 5 μg of protein was loaded into each lane (except the SDS-PAGE lane for BioExpress-N). Protein expression was induced with IPTG in the Celtone and BioExpress media. Labeling code: U = unlabeled (natural abundance); C = carbon-13; N = nitrogen-15; D = deuterium

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Approximate amelogenin yields obtained using various media and induction methods

Media	Induction	Temperature (K)	Approximate yields (mg/ml)
Celtone uniform complete	IPTG	310	10–15
Celtone ^{13}C-, ^{15}N-complete	IPTG	310	10–15
Celtone ^{2}H-, ^{13}C-, ^{15}N-complete	IPTG	310	5–10
Bioexpress ^{15}N, 10x	IPTG	310	10–15
Celtone ^{13}C-, ^{15}N-powder	IPTG	310	~5
Celtone unlabeled powder	IPTG	310	~10
ZY[b]	IPTG	310	~10
ZY	AUTO	310	~30
ZY	AUTO	293	none
LB[a]	AUTO	310	~5
LB	IPTG	310	~15

[a]LB media = 10 g tryptone, 5 g yeast extract, and 10 g NaCl per liter.
[b]ZY media = 10 g tryptone and 5 g yeast extract per liter.

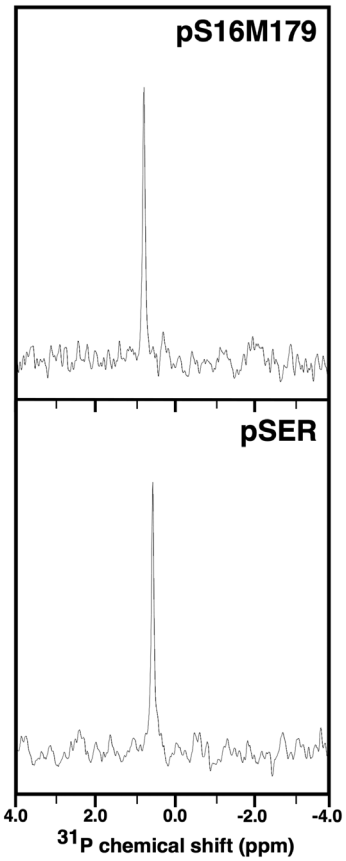


FIGURE 2 One-dimensional ^{31}P NMR spectra of 0.1 mM pS16M179 (top) and 1.0 mM pSer (bottom) collected in 2% acetic acid (pH 2.8) at 293 K and a ^{1}H resonance frequency of 500 MHz. There was no change in both spectra following the incubation of both samples in their NMR tubes for over 1 month at room temperature. The ^{31}P spectra were referenced to 85% phosphoric acid (0.0 ppm).

time frame and, at least for amelogenin expression in *E. coli*, would not require the use of a nonhydrolyzable analog of phosphoserine (Rogerson et al., 2015).

While the one-dimensional ^{31}P spectrum for pS16M179 showed that phosphoserine had been incorporated into M179, we sought additional quantitative information on the homogeneity of amelogenin phosphorylation. Nitrogen-15 labeling provides a sensitive method for quantitating the extent of pSer incorporation through the collection of a "fingerprint" two-dimensional ^{1}H-^{15}N HSQC spectrum. Such fingerprint ^{1}H-^{15}N HSQC spectra have been used to quickly assess the success of expressing ^{15}N-labeled M179 constructs containing point mutations (Buchko et al., 2013; Buchko & Shaw, 2015) or point deletions (Buchko et al., 2018; Tao et al., 2022) by collecting ^{1}H-^{15}N HSQC spectra in 2% acetic acid (pH 2.8), a condition where amelogenin is intrinsically disordered, primarily monomeric in solution (especially at concentrations under 0.5 mM), and the amide chemical shifts have been assigned (Buchko, Bekhazi, et al., 2008; Buchko, Tarasevich, et al., 2008). Figure 3a overlays the ^{1}H-^{15}N HSQC spectrum of M179 (red) and pS16M179 (blue) and shows the disappearance of the Ser-16 amide cross peak in the pS16M179 spectrum and the appearance of a new amide cross peak in this region (black arrows). Analysis of the pS16M179 HNCACB data indicates this amide is connected to a positive and negative cross peak corresponding to $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ with chemical shifts expected for a phosphoserine residue in a disordered protein (Table 2): 2–3 ppm downfield shift for $^{13}C^{\beta}$ and a 1–2 ppm upfield shift for $^{13}C^{\alpha}$ (Bienkiewicz & Lumb, 1999). Further analysis of

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

	Random coil[a]	M179[b]	pS16M179	Ser[c]	pSer[c]
$^{13}C\alpha$	58.1	58.9	57.5	58.6	57.1
$^{13}C\beta$	64.1	63.6	66.2	62.6	65.9
Backbone amide $^{1}H^{N}$	8.36	8.17	8.42	-	-
Backbone amide ^{15}N	116.8	116.0	115.0	-	-
^{31}P	1.2[d]	-	0.7	-	0.5

TABLE 2 Select NMR chemical shifts (ppm) for phosphorylated and non-phosphorylated Ser in M179 and pS16M179 and for the free amino acids pSer and Ser, all in 2% acetic acid, pH 2.8

[a]Serine random coil values are from the BMRB and reflect near neutral pH values.
[b]BMRB entry 15,662 collected on an ~2 mM sample of M179 containing the N-terminal tag MRGSHHHHHHGS-, in 2% acetic acid, pH 2.8 (Buchko, Bekhazi, et al., 2008).
[c]Collected on 1 mM samples under the same conditions as for pS16M179 and M179 (2% acetic acid, pH 2.8).
[d]From Beinkiewicz & Lumb (1999) converted to phosphoric acid referencing.

other backbone assignment data for pS16M179 confirms that this new amide cross peak is preceded by a leucine residue and proceeded by a tyrosine residue, as expected from the primary amino acid sequence for amelogenin in this region, -LSY-. Furthermore, the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled pS16M179 following incubation with lambda phosphatase, an enzyme that removes the phosphate group from phosphorylated serine, threonine, and tyrosine residues (Wera & Hemmings, 1995), was superimposable on the spectrum of non-phosphorylated M179 (data not shown). In conclusion, the NMR data unambiguously shows that pSer has been incorporated into pS16M179 and the new cross peak in its ^{1}H-^{15}N HSQC spectrum is pSer-16.

2.3 | Confirmation of pSer incorporation into pS16M179 by intact mass spectrometry

Analysis of mass spectral data to quantitate the fidelity of phosphoserine incorporation in NMR-isotope labeled proteins is difficult because the isotope labeling is not 100%. Therefore, we used unlabeled Celtone complete media which is prepared identically to ^{13}C-, ^{15}N-labeled Celtone complete media except for ^{13}C- and ^{15}N-enrichment. Moreover, as illustrated in Figure 1, the level of unphosphorylated amelogenin appears similar using unlabeled and ^{13}C-, ^{15}N-labeled Celtone complete media and Table 2 shows the yield of amelogenin is also similar. Figure 4 illustrates the deconvoluted monoisotopic mass spectrum for pS16M179 obtained from unlabeled Celtone complete media. The major peak at *m/z* 20,229.3 corresponds closely to the expected molecular weight of pS16M179 (20,228.1 Da). To the right of this major peak is a series of *m/z* peaks that increase by ~15 *m/z* units each. These are likely due to the oxidation of methionine residues which occurs spontaneously during the aqueous storage of protein under oxic conditions (Bettinger et al., 2020; Glaser & Li, 1974) (M179 contains nine methionine residues). Out of all the other minor peaks in the spectrum the only one corresponding to the substitution of a native amino acid for pSer is at *m/z* 20,149.3 which corresponds to a Ser residue. Because no other native amino acid substitutes for pSer during translation, this Ser likely is a result of the dephosphorylation of pSer inside *E. coli* cells prior to cell harvesting (centrifugation and then freezing at 193 K). It is unlikely due to dephosphorylation during the purification protocol because all the purifications were done similarly and included heating at 343 K under acidic conditions as the first step which will likely quickly denature any phosphatases. Note that this could also explain the higher level of non-phosphorylated M179 using Celtone-CDN (Figure 3) as these cells were harvested ~16 h after induction instead of ~4 h. Hence, the fidelity of pSer incorporation into M179 is likely near 100% but the final lower yields are due to dephosphorylation after translation. Assuming little difference in the ability to ionize M179 and pS16M179 in the gas phase during mass spectrometry data collection, the level of non-phosphorylated amelogenin after purification was therefore ~5% using unlabeled Celtone media. Given that the Phos-tag gel in Figure 1 suggests the level of unphosphorylated amelogenin is similar using either unlabeled or ^{13}C-, ^{15}N-labeled Celtone complete media, ~95% of the ^{13}C-, ^{15}N-labeled sample is phosphorylated at Ser-16 which agrees with the ^{1}H-^{15}N HSQC data. That is, there was no evidence for the Ser-16 amide cross peak above the noise in the ^{1}H-^{15}N HSQC spectrum for pS16M179 although it is evident above the noise in the spectrum from Celtone-CDN where the level of non-phosphorylated amelogenin is ~15% judging from the Phos-tag gel in Figure 3 (data not shown).

While analysis of the two- and three-dimensional NMR data convincingly shows that pS16M179 is phosphorylated at the correct position of the sequence, this was further corroborated by the analysis of fragmentation patterns for the major species in the mass spectrum of pS16M179 in Figure 4 as described in Figure S2.

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

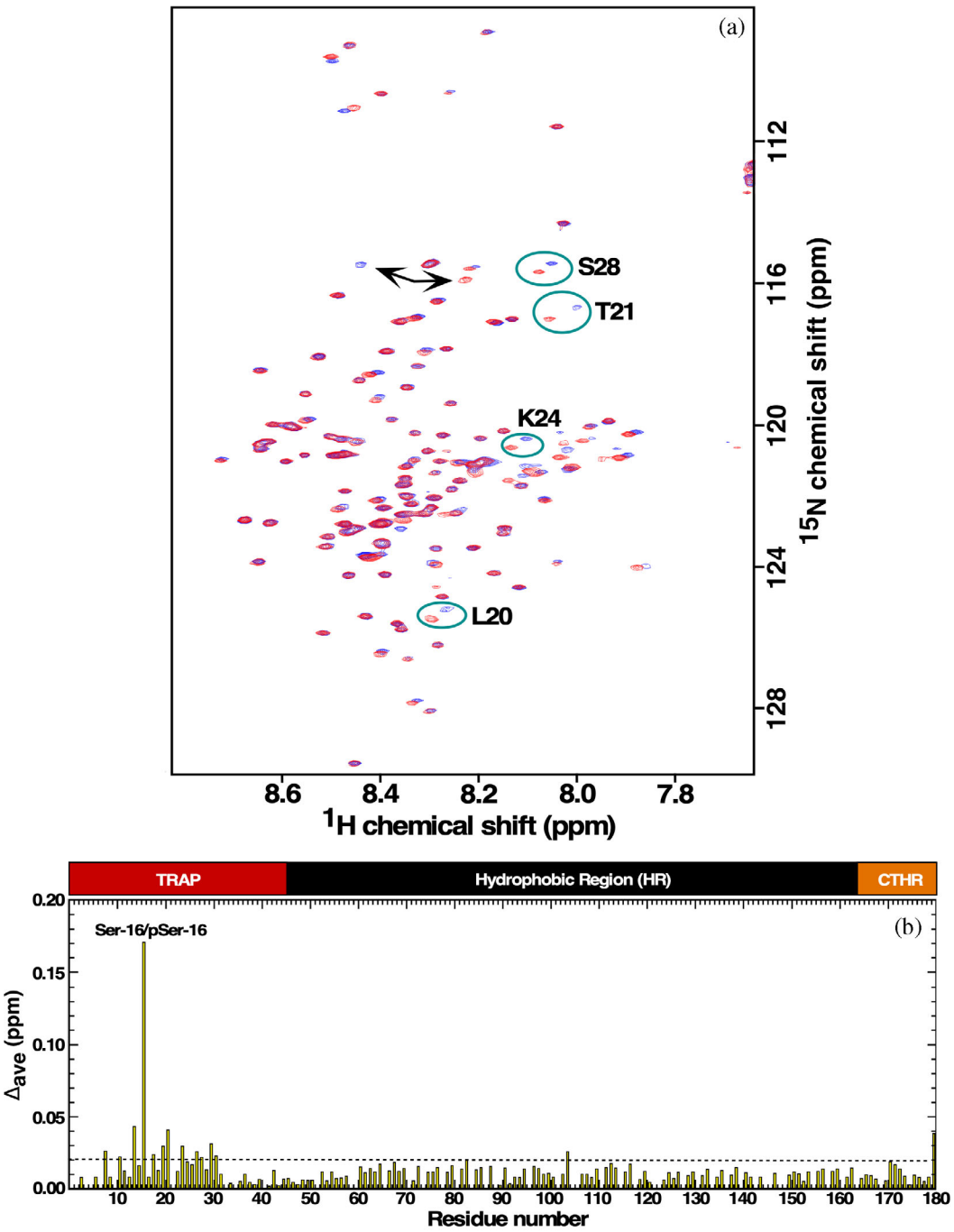


FIGURE 3 (a) Overlay of the ^{1}H-^{15}N HSQC spectra for ^{13}C-,^{15}N-labeled M179 (red) upon pS16M179 (blue) collected in 2% acetic acid (pH 2.8) at 293 K. Spectra were collected on 0.3 mM samples, at a ^{1}H resonance frequency of 600 MHz. (b) Plot of the average chemical shift change, Δ_{ave} ppm $= \{[(\Delta^{1}H^{N})^{2} + (\Delta^{15}N/5)^{2}]/2\}^{1/2}$, between the amide resonances of pS16M179 and M179 extracted from ^{1}H-^{15}N HSQC spectra. The most perturbed resonance is the chemical shift difference between Ser-16 and pSer-16 (labeled with arrows) with mild perturbations confined to a region surrounding pSer-16 in the N-terminal. On top of the plot is a cartoon representation of the three regions present in amelogenin: N-terminal tyrosine-rich region (TRAP; red), hydrophobic region (HR; black) and C-terminal hydrophilic region (CTHR; orange)

2.4 | Backbone amide chemical shift perturbations between pS16M179 and M179

The average combined perturbations between the ^{1}H-^{15}N HSQC spectra of pS16M179 and M179 are shown in Figure 3b. The plot indicates that most of the chemical shift perturbations are less than 0.02 ppm (dashed line), a small value that is likely due to the resolution limits of the data (~6 Hz/point ^{1}H, ~12 Hz/point ^{15}N) or perhaps minor differences in the concentration of both samples.

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

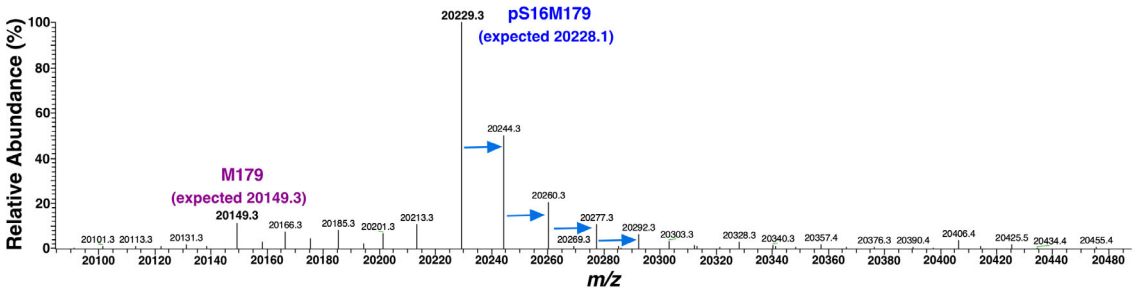


FIGURE 4 Deconvoluted monoisotopic intact mass spectrum for the unlabeled Celtone product (Celtone-U) shown in the Figure 3 gels. The most abundant species at $m/z = 20{,}229.3$ corresponds to pS16M179 with an expected molecular weight of 20,228.1 Da. The four periodic species to the right of the major peak (highlighted with blue arrows) are protein oxidation products of increasing size, likely due to methionine oxidation. Out of the minor peaks, the only one consistent with the substitution of a native amino acid for pSer during translation is at $m/z = 20{,}149.3$ which corresponds to native M179 containing a serine with an expected molecular weight of 20,149.3 Da. The absence of peaks due to the substitution of any of the other 19 native amino acids suggests that pSer incorporation likely approaches 100% during translation and is dephosphorylated *in vivo* prior to purification

The most perturbed resonance is for Ser-16/pSer-16, with an average delta ppm of approximately 0.17 ppm. Around this major perturbation are a small cluster of resonances between Gly-8 to Arg-31, however, none of these perturbations are greater than 0.05 ppm. Previous site specific substitutions in M179, Thr-21 to Ile, and Pro-41 to Thr, generated larger chemical shift perturbations about the site of the substitution, up to ~0.3 ppm, under similar conditions of pH (Buchko et al., 2013). Hence, in 2% acetic acid at pH 2.8 where amelogenin is monomeric and intrinsically disordered, the substitution of pSer for Ser in pS16M179 had no significant structural effect on murine amelogenin (aside from Ser-16/pSer-16, there were no significant changes in any of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for the other residues) nor significant perturbations to the local electronic environment.

2.5 | ^{31}P NMR spectra of pS16M179 in different quaternary states

At pH values above approximately seven amelogenin self-assembles into nanospheres composed of 20–100 monomers (Du et al., 2005; Fang et al., 2011; Fincham et al., 1995). These structures are believed to be essential for proper enamel formation with the N-terminal region identified as playing an essential role in amelogenesis (Paine & Snead, 1997). Single tryptophan fluorescence (constructs with only a single tryptophan, Trp-25, Trp-45, or Trp-161) (Bromley et al., 2011) and NMR studies (Buchko et al., 2022) suggest the N-terminal is buried within the nanospheres. This is corroborated by the absence of any significant ^{31}P resonances in the one-dimensional ^{31}P NMR spectrum for 0.5 mM pS16M179 at pH 7.1 (Figure 5 (top)) after 20 h of data collection. The absence of a major ^{31}P resonance is likely due to the same

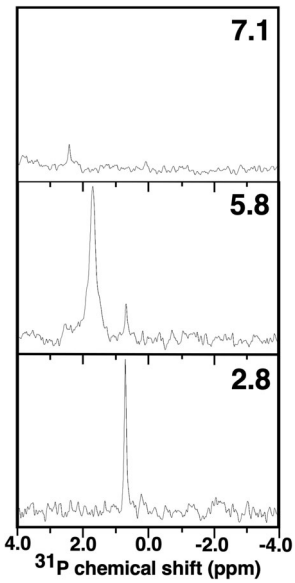


FIGURE 5 One-dimensional ^{31}P NMR spectra of pS16M179 at pH 7.1 (top), pH 5.8 (middle) and pH 2.8 (bottom) collected at 293 K and a ^{1}H resonance frequency of 500 MHz. The top two spectra were for a 0.5 mM sample and the bottom spectrum is a 0.1 mM sample. The ^{31}P spectra were referenced to 85% phosphoric acid (0.0 ppm)

reasons amide resonances were not observed for the entire N-terminal region of M179 in ^{1}H-^{15}N HSQC spectra collected on ^{2}H-, ^{13}C-, ^{15}N-labeled M179 nanospheres (Buchko et al., 2022). In these ^{1}H-^{15}N HSQC spectra the disappearance or intensity reduction of amide resonances were interpreted to reflect changes in dynamics (intermediate ms-μs motion) and/or multiple chemical environments (heterogenous interfaces) of amide nuclei at protein–protein interfaces (Buchko, Tarasevich, et al., 2008). While the absence of a major ^{31}P resonance is consistent with the N-terminal region around pSer-16

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

being buried and at protein-protein interfaces, it is not buried 100% of the time. The observation that lambda phosphatase was able to remove the phosphate group from pS16M179 in buffers favorable for nanosphere formation (pH 7.1 and pH 7.8), as shown in Figure 1, is in agreement with previous studies suggesting monomers (Fincham et al., 1994; Limeback & Simic, 1990) or small oligomers (Tao et al., 2015) are in dynamic equilibrium with nanospheres. Indeed, there is a minor, relatively sharp resonance at 2.2 ppm which may represent the pSer-16 resonance in monomeric amelogenin that was observed at 1.2 ppm at pH 2.8 (Table 1 and Figure 4). This is consistent with the pH dependence of the ^{31}P chemical shift over the pH range of 2–9, moving downfield with increasing pH (Bienkiewicz & Lumb, 1999; Hoffmann et al., 1994).

An attempt was made to follow the lambda phosphatase dephosphorylation of pS16M179 by ^{31}P NMR spectrometry using an unlabeled sample (0.5 mM). As just noted, no major ^{31}P resonance was observed in the one-dimensional spectrum before the additional of the phosphatase (Figure 4, top), but, it was anticipated that it would be possible to follow the reaction by the appearance of a signal for the free phosphate. Although lambda phosphatase had converted pS16M179 completely to M179, as shown in the Phos-tag gel in Figure 1, no signal was observed for the free phosphate in the one-dimensional ^{31}P NMR spectrum. This suggests that after cleavage the free phosphate bound to amelogenin nanospheres.

While the N-termini of amelogenin appears to be largely buried in nanospheres at pH values above seven, this region appears less buried in the transition from monomers to oligomers as the pH is increased from 3 to ~6.8 (Bromley et al., 2011). A chemical shift perturbation study as a function of pH enabled the tracking of the protein–protein interface at the residue level and showed that at pH 6.6 Ser-16 and a region around it, Ser-9 to Leu-20, was not intimately part of an oligomer interface (Buchko et al., 2022). This is corroborated in Figure 5 (middle), the one dimensional ^{31}P NMR spectrum for 0.5 mM pS16M179 at pH 5.8, where a major resonance is observed at 1.7 ppm. Not only is this resonance shifted relative to the single resonance observed at pH 2.8 (Figure 5, bottom), it is also broader, a characteristic of a species with a larger molecular weight. Note that a second minor resonance at 0.7 ppm, a similar chemical shift observed for pS16M179 at pH 2.8, is also present. While it would be tempting to suggest this is the chemical shift for monomeric pS16M179, the pKa of the phosphoryl group of pSer in peptides is 5.96 and the ^{31}P chemical shift is pH dependent over the pH range of 2–9, moving downfield with increasing pH (Bienkiewicz & Lumb, 1999; Hoffmann et al., 1994). Hence, the reason for the two resonances is unclear especially since only a single resonance was observed for the amide chemical shift of S16 in M179 at pH 6.6 (Buchko et al., 2022). Efforts are in progress to determine the reason for the two ^{31}P resonances for pS16M179, including the collection of NMR backbone assignment data on ^{2}H-, ^{13}C-, ^{15}N-labeled pS16M179 at pH 5.8.

3 | CONCLUSIONS

Using third generation genetic code expansion protocols we have shown it is possible to express and purify recombinant, unlabeled, bacterial murine amelogenin in high yields with a level of pSer incorporation at the native Ser-16 position approaching 95% (the fidelity of incorporation is likely near 100% but some *in vivo* dephosphorylation after translation is suspected). Using enriched, ^{13}C- and ^{15}N-labeled commercial media it is possible to uniformly ^{13}C-, ^{15}N-label amelogenin for NMR studies with the same level of pSer incorporation. Sacrificing yield (~50% less, Table 1) and pSer incorporation levels (~85%, Figure 2), amelogenin may be perdeuterated for NMR experiments designed to assist studies of large amelogenin quaternary structures (Buchko et al., 2022) using enriched, ^{2}H-, ^{13}C-, and ^{15}N-labeled commercial media. The approaches for NMR labeling all use IPTG to induce protein expression. For experiments not requiring NMR isotope labeling of amelogenin and only pSer incorporation, ZY- or LB-media may be used instead of unlabeled enriched commercial media (Celtone or BioXpress) with IPTG induction of protein expression to obtain various levels (~85%–95%) of pSer incorporation (as estimated by the Phos-tag gel in Figure 1). While our final levels of pSer incorporation were high, certainly above levels where they would not complicate interpretation of NMR data, they were not 100%. For pS16M179 experiments requiring near 100% levels of pSer incorporation, we suspect that it may be possible to coax the *E. coli* closer toward 100% by optimizing growth conditions yet further through experimenting with growth parameters such as: optical density at induction, the concentration of IPTG used at induction, the temperature, and perhaps most importantly, the length of induction before cell harvesting as mass spectrometry suggests non-phosphorylated M179 is largely due to dephosphorylation of pS16M179 in *E. coli* after translation.

The observation of two resonances in the one-dimensional ^{31}P NMR spectrum of pS16M179 at pH 5.8 is just one example of the new insights to be discovered regarding the biophysical properties of amelogenin provided by the ability to site-specifically incorporate pSer

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

into the primary amino acid sequence of amelogenin. Moreover, because the major NMR-detectable natural isotope of phosphorus, phosphorus-31, is 100% abundant, unlabeled pS16M179 contains a label near the N-terminal that can be readily followed by ^{31}P NMR spectroscopy. In conclusion, the high-yield ^{13}C-, ^{15}N-labeling of amelogenin with high levels of phosphoserine incorporation shown here will not only accelerate biomineralization research to understand amelogenesis, but, trigger the expanded use of the third generation genetic code expansion protocols to introduce phosphorylated amino acids into other proteins to better understand the structural and functional consequences of the most common post-translational modification cells use to regulate protein function.

4 | METHODS

4.1 | Cloning, expression, and purification

DNA encoding the primary amino acid sequence for native murine amelogenin (Simmer et al., 1994) was codon optimized with a TAG codon inserted for Ser-16, synthesized by IDT (Corvaville, Iowa), and inserted into a pRBC plasmid (Addgene # 173897) containing an ampicillin resistance gene (Zhu et al., 2019) using PPY-based SLiCE techniques (Zhang et al., 2012). The plasmid containing the machinery for pSer incorporation, pKW2-EPSep, contained a chloramphenicol resistance gene and was a gift from Jason Chin. Initial evaluations for expressing phosphorylated sfGFP (sfGFP-150TAG) in the release Factor-1 (RF1) deficient *E. coli* B95(DE3) ΔA ΔfabR ΔserB strain in media amenable to isotopic enrichment (Figure S1) were performed as previously described (Vesely et al., 2022). Subsequent protein expression was achieved following a 4 day protocol (Zhu et al., 2019). *Day 1*. Both plasmids were used to simultaneously transform a competent, Release Factor-1 (RF1) deficient, *E. coli* B95(DE3) ΔA ΔfabR ΔserB strain using a standard heat shock method (30 min ice; 315 K, 45 s; 2 min ice). After 1 ml addition of SOC media, the cells were allowed to recover with shaking for 2 h at 310 K. All the cells were streaked on ampicillin (25 μg/ml) and chloramphenicol (7 μg/ml) LB-agar plates by gently spinning (3 min @ 3000 g), removing top 900 μl, and resuspending the pellet in the remaining 100 μl of media. *Day 2*. Following overnight incubation at 310 K, the plates were transferred to a 277 K fridge for the day. Late afternoon, a 25 ml starter culture was prepared starting with approximately 10 colonies using a buffered, glucose-rich, non-autoinducing medium (ZY-NIM) (Zhu et al., 2019) containing ampicillin (50 μg/ml) and chloramphenicol (14 μg/ml). The culture was shaken at 310 K overnight in a baffled 250 ml flask at ∼200 rpm. *Day 3*. The optical density of the starter culture at 600 nm was measured to ensure it had grown to saturation (OD_{600} = 3–8). Upon reaching growth saturation, 1 ml of this culture was added to 100 ml of ^{13}C-, ^{15}N-labeled Celtone complete media (Cambridge Isotopes Laboratory, Tewksbury, MA) in a 500 ml baffled flask and incubated at 310 K with shaking (200 rpm). At an OD_{600} of ∼0.8 protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.026 mg/ml) and incubation continued under the same conditions. *Day 4*. After approximately 4 h the cells (OD_{600} ∼ 1.7) were harvested by mild centrifugation and frozen at 193 K.

In addition to using 100 ml of ^{13}C-, ^{15}N-labeled Celtone complete media on *Day 3* in the protein expression protocol, other commercial media were also tested including ^{2}H-, ^{13}C-, ^{15}N-labeled and unlabeled Celtone complete media, ^{15}N-labeled BioExpress complete media (10× concentrate; Cambridge Isotopes Laboratory, Tewksbury, MA), and ^{13}C-, ^{15}N-labeled and unlabeled Celtone powder (Cambridge Isotopes Laboratory, Tewksbury, MA) (0.2% w/v in Miller minimal media). Non-commercial media tested were LB and ZY using both IPTG induction (0.026 mg/ml at OD_{600} of ∼0.8) and auto-induction (Studier, 2005). Cells were harvest ∼4 h after induction with IPTG except with ^{2}H-, ^{13}C-, ^{15}N-labeled Celtone complete media where harvesting occurred ∼16 h later. Harvesting of autoinduction cultures occurred the following morning.

Amelogenin was purified following a previously published protocol (Buchko & Shaw, 2015). For 100 ml of cell culture, 20 ml of 2% acetic acid was added to the frozen pellet in a 50 ml centrifuge tube. The solution was incubated at ∼343 K for ∼15 min in a water bath (with a couple manual hard shakes in between), sonicated for 2 min (5 s on, 2 s off, 25% level) on a Qsonica GEX750 sonicator (Newton, CT), incubated for a further 15 minutes in the water bath followed by another 2 min of sonication under the same conditions. The entire solution was placed in a 3.5 kDa molecular weight cutoff dialysis tubing (Fisher Scientific, Waltham, MA) and dialyzed in 5 l of 2% aqueous acetic acid twice (over 24–36 h) at 277 K. The insoluble cell debris was removed by centrifugation for 45 min at 17,500 rpm in a Beckman J-20 rotor in a Beckman Avanti J-25 centrifuge (Fullerton, CA). The supernatant was applied to a Resource-RPC 3 ml column (GE Healthcare, Uppsala, Sweden) in 5–10 ml volumes and the following linear purification gradient applied: Buffer A = 100% water in 0.005% TFA, Buffer B = 70% aqueous acetonitrile in 0.005% TFA; flow rate = 3 ml/min; Step 1–100% Buffer A 5 CV, Step 2––

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

linear gradient 0–100% Buffer B. The fractions containing amelogenin eluting at ~55% Buffer B were pooled, frozen at 193 K, and lyophilized. To assess the purity of the products and the extent of pSer incorporation, regular 4%–12% TrisHCl SDS-PAGE gels and 12.5% SuperSep Phos-tag SDS-PAGE gels (199–18,011, FUJIFILM Wako Chemicals USA, Richmond, VA), respectively, were run. The Phos-tag gel was run following the manufacturer's instructions at 120 V constant voltage for 2 h and stained with GelCode Blue Safe Protein Stain (ThermoFisher Scientific, Rockford, IL). The yields of purified amelogenin were as measured using absorption at 280 nm, a calculated molar extinction coefficient of 25,440 M^{-1} cm^{-1}, and a molecular weight of 20,160 Da. Note that amelogenin yields (total protein) were ~50% greater if 6 M guanidinium hydrochloride was used instead of 2% acetic acid in the cell lysis step (followed by centrifugation prior to dialysis) (Buchko & Shaw, 2015). However, this method resulted in some additional impurities that co-eluted with amelogenin on the reverse phase column, and therefore, this method was avoided in these studies.

4.2 | NMR spectroscopy

Lyophilized pS16M179 was resuspended in an NMR buffer consisting of 2% CD_3CO_2D, 7% D_2O/93% H_20, pH 2.8, to a concentration of 0.3 mM (measured using a calculated ε of 25,440 M^{-1} cm^{-1}). All the NMR data was collected at 293 K on four-channel Varian-600 NMR spectrometers equipped with triple-resonance probes and pulse field gradients to verify chemical shift assignment and the presence of pSer in the sequence. This included two- and three-dimensional ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, HNCA, HNCO, HNCACB, HNCOCACB, and HNN experiments (Varian Biopack pulse programs). One-dimensional ^{31}P NMR spectra were collected for pS16M179 and pSer on a Agilient-DD2 NMR spectrometer operating at a ^{1}H proton resonance frequency of 499.92 MHz with proton decoupling and referenced to 85% phosphoric acid (0.0 ppm). The O-phospho-L-serine (P0878) was purchased from Sigma-Aldrich (St. Louis, MO).

4.3 | Intact protein mass spectrometry

The protein solution at 1.2 μg/μl was diluted in water to 24 ng/μl. Ten μl was loaded (~240 ng) onto a Waters NanoAcquity LC with online desalting. Reversed phase separation was carried out on an in-house packed C2 column (100 μm i.d., ~50 cm long, packing material SMTC2MEB2–3-300 from Separation Methods Technologies, Newark, DE). Mobile phases were 0.2% formic acid in water (A) and 0.2% formic acid in acetonitrile (B). A linear gradient with a flow rate of 0.3 μl/min was run from 15% to 50% mobile phase B over 30 min. The protein eluted between 30%–40% mobile phase B. Mass spectrometry data were collected on a Thermo Orbitrap Exploris. The mass spectrum (120 k resolution, 5 microscans) across the elution window were summed, and deconvoluted with FreeStyle (v1.5, ThermoScientific) for estimating the relative abundance of the phosphorylated protein.

4.4 | Lambda phosphatase reaction

Approximately 0.8 mg of lyophilized, ^{15}N-labeled pS16M179 was resuspended in 250 μl of water and allowed to sit at room temperature for 24 h. This solution was added to a 250 μl solution containing 50 mM TrisHCl, pH 8.5, while gently vortexing. Fifty μl of 10× Reaction Buffer (500 mM HEPES, 1 mM EGTA, 50 mM dithiothreiotol, 0.1% BRIJ 35), 50 μl of 10× $MgCl_2$ (20 mM), and 5 μl of lambda phosphatase (400 units/uL, Santa Cruz Biotechnology, Dallas, TX) was added to this solution that had a final pH of 7.8. The reaction solution was incubated overnight at 303 K. Prior to purification by reverse phase chromatography as described above, an equal volume of reverse phase Buffer A was added to the solution and the pH adjusted to ~3 with 1 μl additions of TFA. A similar experiment was performed on a 0.5 mM sample of unlabeled pS16M179 in a Shigemi NMR tube (260 μl) under the same buffer conditions except the pH was adjusted to 7.1 prior to the addition of 5 μl of lambda phosphatase.

AUTHOR CONTRIBUTIONS

Garry W. Buchko: Conceptualization (lead); data curation (lead); funding acquisition (equal); investigation (lead); methodology (equal); resources (equal); supervision (lead); visualization (lead); writing - original draft (lead); writing – review and editing (lead). **Mowei Zhou:** Investigation (supporting). **Cat Hoang Vesely:** Investigation (supporting); methodology (supporting). **Jinhui Tao:** Funding acquisition (equal); resources (equal). **Wendy J. Shaw:** Funding acquisition (lead); project administration (lead). **Ryan A. Mehl:** Conceptualization (supporting); investigation (supporting); methodology (equal); resources (supporting); writing – review and editing (supporting). **Richard B. Cooley:** Conceptualization (supporting); investigation (supporting); methodology (equal); resources (supporting); supervision (supporting); visualization (supporting); writing – review and editing (supporting).

ACKNOWLEDGMENTS

This research was supported by NIH-NIDCR Grant number DE-015347 and performed at the Pacific Northwest

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

National Laboratory (PNNL), a facility operated by Battelle for the U.S. Department of Energy, including access to the W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the U.S. DOE Biological and Environmental Research program. Battelle operates PNNL for the U.S. Department of Energy under contract DE-AC05-76RL01830. Additional support was provided by the GCE4All Biomedical Technology Development and Dissemination Center supported by National Institute of General Medical Science grant RM1-GM14422.

DATA AVAILABILITY STATEMENT

Data available upon request from the authors.

ORCID

Garry W. Buchko https://orcid.org/0000-0002-3639-1061
Mowei Zhou https://orcid.org/0000-0003-3575-3224
Wendy J. Shaw https://orcid.org/0000-0002-4696-7415

REFERENCES

Bettinger JQ, Welle KA, Hryhorenko JR, Ghaemmaghami S. Quantitative analysis of *in vivo* methionine oxidation of the human proteome. J Proteome Res. 2020;19:624–33.

Beyer JN, Hosseinzadeh P, Gottfried-Lee I, Van Fossen EM, Zhu P, Bednar RM, et al. Overcoming near-cognate suppression in a release factor 1-deficient host with an improved nitro-tyrosine tRNA synthetase. J Mol Biol. 2020;432:4690–704.

Bienkiewicz EA, Lumb KJ. Random-coil chemical shifts of phosphorylated amino acids. J Biomol NMR. 1999;15:203–6.

Bromley KM, Kiss AS, Lokappa SB, Lakshminarayanan R, Fan D, Ndao M, et al. Dissecting amelogenin protein nanospheres: characterization of metastable oligomers. J Biol Chem. 2011;286:34643–53.

Buchko GW, Bekhazi J, Cort JR, Valentine NB, Snead ML, Shaw WJ. ^{1}H, ^{13}C, and ^{15}N resonance assignments of murine amelogenin, an enamel biomineralization protein. Biomol NMR Assign. 2008;2:89–91.

Buchko GW, Jayasinha Arachchige R, Tao J, Tarasevich BJ, Shaw WJ. Identification of major matrix metalloproteinase-20 proteolytic processing products of murine amelogenin and tyrosine-rich amelogenin peptide using a nuclear magnetic resonance spectroscopy based method. Arch Oral Biol. 2018;93:187–94.

Buchko GW, Lin G, Tarasevich BJ, Shaw WJ. A solution NMR investigation into the impaired self-assembly properties of two murine amelogenins containing the point mutations T21I or P41T. Arch Biochem Biophys. 2013;537:217–24.

Buchko GW, Mergelsberg ST, Tarasevich BJ, Shaw WJ. Residue specific insights into the protein-protein interfaces driving amelogenin self-assembly in solution. Biochemistry. 2022;61:2909–21.

Buchko GW, Shaw WJ. Improved protocol to purify untagged amelogenin—application to murine amelogenin containing the equivalent P70-->T point mutation observed in human *amelogenesis imperfecta*. Protein Expr Purif. 2015;105:14–22.

Buchko GW, Tarasevich BJ, Bekhazi J, Snead ML, Shaw WJ. A solution NMR investigation into the early events of amelogenin nanosphere self-assembly initiated with sodium chloride or calcium chloride. Biochemistry. 2008;47:13215–22.

Carneiro KMM, Zhai HL, Zhu L, Horst JA, Sitlin M, Nguyen M, et al. Amyloid-like ribbons of amelogenins in enamel mineralization. Sci Rep. 2016;6:6.

Daculsi G, Menanteau J, Kerebel LM, Mitre D. Length and shape of enamel crystals. Calcif Tissue Int. 1984;36:550–5.

Delak K, Harcup C, Lakshminarayanan R, Sun Z, Fan Y, Moradian-Oldak J, et al. The tooth enamel protein, procine amelogenin, is an intrinsically disordered protein with an extended molecular configuration in the monomeric form. Biochemistry. 2009;48:2272–81.

Du C, Falini G, Fermani S, Abbott C, Moradian-Oldak J. Supramolecular assembly of amelogenin nanospheres into birefringent microribbons. Science. 2005;307:1450–4.

Engelberth SA, Bacino MS, Sandhu S, Li W, Bonde J, Habelitz S. Progression of self-assembly of amelogenin protein supramolecular structures in simulated enamel fluid. Biomacromolecules. 2018;19:3917–24.

Fang PA, Conway JF, Margolis HC, Simmer JP, Beniash E. Hierarchical self-assembly of amelogenin and the regulation of biomineralization at the nanoscale. Proc Natl Acad Sci U S A. 2011;108:14097–102.

Fang PA, Margolis HC, Conway JF, Simmer JP, Beniash E. Cryo-TEM study of effects of phosphorylation on the hierarchical assembly of porcine amelogenin and its regulation of mineralization *in vitro*. J Struct Biol. 2013;183:250–7.

Fincham AG, Moradian-Oldak J, Diekwisch TGH, Lyaruu DM, Wright JT, Bringas P Jr, et al. Evidence for amelogenein "nanospheres" as functional components of secretory-stage enamel matrix. J Struct Biol. 1995;115:50–9.

Fincham AG, Moradianoldak J, Simmer JP, Sarte P, Lau EC, Diekwisch T, et al. Self-assembly of a recombinant amelogenin protein generates supramolecular structures. J Struct Biol. 1994;112:103–9.

Gibson CW, Yaun ZA, Hall B, Longenecker G, Chen EH, Thyagarajan T, et al. Amelogenin-deficient mice display an *amelogenesis imperfecta* type phenotype. J Biol Chem. 2001;276:31871–5.

Glaser CB, Li CH. Reaction of bovine growth-hormone with hydrogen-peroxide. Biochemistry. 1974;13:1044–7.

He XD, Wu SP, Martinez-Avila O, Cheng YF, Habelitz S. Self-aligning amelogenin nanoribbons in oil-water system. J Struct Biol. 2011;174:203–12.

Hoffmann R, Reichert I, Wachs WO, Zeppezauer M, Kalbitzer HR. H-1 and P-31 NMR-spectroscopy of phosphorylated model peptides. Int J Pept Prot Res. 1994;44:193–8.

Hu Y, Smith CE, Cai Z, Donnelly LA, Yang J, Hu JC, et al. Enamel ribbons, surface nodules, and octacalcium phosphate in C57BL/6 Amelx(−/−) mice and Amelx(+/−) lyonization. Mol Genet Genomic Med. 2016;4:641–61.

Hunter GK. Interfacial aspects of biomineralization. Curr Opin Solid State Mat Sci. 1996;1:430–5.

Kinoshita E, Kinoshita-Kikuta E, Takiyama K, Koike T. Phosphate-binding tag, a new tool to visualize phosphorylated proteins. Mol Cell Proteomics. 2006;5:749–57.

Kwak SY, Wiedemann-Bidlack FB, Beniash E, Yamakoshi Y, Simmer JP, Litman A, et al. Role of 20-kDa amelogenin (P148)

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

phosphorylation in calcium phosphate formation *in vitro*. J Biol Chem. 2009;284:18972–9.
Lakshminarayanan R, Fan D, Du C, Moradian-Oldak J. The role of secondary structure in the entropically driven amelogenin self-assembly. Biophys J. 2007;93:3664–74.
Limeback H, Simic A. Biochemical-characterization of stable high-molecular-weight aggregates of amelogenins formed during porcine enamel development. Arch Oral Biol. 1990;35:459–68.
Lu JX, Xu YS, Shaw WJ. Phosphorylation and ionic strength alter the LRAP-HAP interface in the N-terminus. Biochemistry. 2013;52:2196–205.
Margolis HC, Beniash E, Fowler CE. Role of macromolecular assembly of enamel matrix proteins in enamel formation. J Dent Res. 2006;85:775–93.
Moradian-Oldak J, Leung W, Fincham AG. Temperature and pH-dependence of amelogin self-assembly: a particle size distribution study. J Struct Biol. 1998;122:320–7.
Moradian-Oldak J, Simmer JP, Lau EC, Sarte PE, Slavkin HC, Fincham AG. Detection of monodisperse aggregates of a recombinant amelogenin by dynamic light scattering. Biopolymers. 1994;34:1339–47.
Opitz C, Ahrne E, Goldie KN, Schmidt A, Grzesiek S. Deuterium induces a distinctive *Escherichia coli* proteome that correlates with the reduction in growth rate. J Biol Chem. 2019;294:2279–92.
Paine ML, Snead ML. Protein interactions during assembly of the enamel organic extracellular matrix. J Bone Min Res. 1997;12:221–7.
Paine ML, Zhu DH, Luo W, Snead ML. The amelogenin phenotype confirmed in transgenic animals. J Dent Res. 2000;79:513–3.
Pawson T, Scott JD. Protein phosphorylation in signaling—50 years and counting. Trends Biochem Sci. 2005;30:286–90.
Ravassipour D, Hart PS, Hart TC, Ritter AV, Yamauchi M, Gibson C, et al. Unique enamel phenotype associated with amelogenin gene (AMELX) codon 41 point mutation. J Dent Res. 2000;79:1476–81.
Rogerson DT, Sachdeva A, Wang KH, Haq T, Kazlauskaite A, Hancock SM, et al. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat Chem Biol. 2015;11:496–503.
Shaw WJ, Tarasevich BJ, Buchko GW, Arachchige RMJ, Burton SD. Controls of nature: secondary, tertiary, and quaternary structure of the enamel protein amelogenin in solution and on hydroxyapatite. J Struct Biol. 2020;212:107630.
Shin NY, Yamazaki H, Beniash E, Yang X, Margolis SS, Pugach MK, et al. Amelogenin phosphorylation regulates tooth enamel formation by stabilizing a transient amorphous mineral precursor. J Biol Chem. 2020;295:1943–59.
Simmer JP, Fincham AG. Molecular mechanisms of dental enamel formation. Crit Rev Oral Biol Med. 1995;6:84–108.
Simmer JP, Lau EC, Hu CC, Aoba T, Lacey M, Nelson D, et al. Isolation and characterization of a mouse amelogenin expressed in *Escherichia coli*. Calcif Tissue Int. 1994;54:312–9.
Smith CEL, Poulter JA, Antanaviciute A, Kirkham J, Brookes SJ, Inglehearn CF, et al. *Amelogenesis imperfecta*; genes, proteins, and pathways. Front Physiol. 2017;8:435.
Studier WF. Production of auto-induction in high-density shaking cultures. Protein Expr Purif. 2005;41:207–34.
Tao J, Hansen E, Dohnalkova A, Buchko GW, Jin B, Shaw WJ, et al. Changes in the C-terminal, N-terminal, and histidine regions of amelogenin reveal the role of oligomer structure on adsorption and hydroxyapatite mineralization. Front Physiol. 2022;13:1034662.
Tao JH, Buchko GW, Shaw WJ, De Yoreo JJ, Tarasevich BJ. Sequence-defined energetic shifts control the disassembly kinetics and microstructure of amelogenin adsorbed onto hydroxyapatite (100). Langmuir. 2015;31:10451–60.
Ten Cate AR. Oral histology: developement, structure, and function. St. Louis: Mosby; 1994.
Toyosawa S, OhUigin F, Figueroa F, Tichy H, Klein J. Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. Proc Natl Acad Sci U S A. 1998;95:13056–61.
Ubersax JA, Ferrell JE. Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Bio. 2007;8:530–41.
Vesely CH, Reardon PN, Yu Z, Barber E, Mehl RA, Cooley RB. Accessing isotopically labeled proteins containing genetically encoded phosphoserine for NMR with optimized expression conditions. J Biol Chem. 2022;298:102613.
Wera S, Hemmings BA. Serine threonine protein phosphatases. Biochem J. 1995;311:17–29.
Wiedemann-Bidlack FB, Beniash E, Yamakoshi Y, Simmer JP, Margolis HC. PH triggered self-assembly of native and recombinant amelogenins under physiological pH and termperature *in vitro*. J Struct Biol. 2007;160:57–69.
Wiedemann-Bidlack FB, Kwak SY, Beniash E, Yamakoshi Y, Simmer JP, Margolis HC. Effects of phosphorylation on the self-assembly of native full-length porcine amelogenin and its regulation of calcium phosphate formation *in vitro*. J Struct Biol. 2011;173:250–60.
Witkop CJ Jr, Kuhlmann W, Sauk J. Autosomal recessive pigmented hypomaturaton a*melogenesis imperfecta*: report of a kindred. Oral Surg Oral Med Oral Pathol. 1973;36:367–82.
Yamazaki H, Beniash E, Yamakoshi Y, Simmer JP, Margolis HC. Protein phosphorylation and mineral binding affect the secondary structure of the leucine-rich amelogenin peptide. Front Physiol. 2017;8:8.
Zhang YW, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. 2012;40:e55.
Zhu P, Gafken PR, Mehl RA, Cooley RB. A highly versatile expression system for the production of multiply phosphorylated proteins. ACS Chem Biol. 2019;14:1564–72.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Buchko GW, Zhou M, Vesely CH, Tao J, Shaw WJ, Mehl RA, et al. High-yield recombinant bacterial expression of ^{13}C-, ^{15}N-labeled, serine-16 phosphorylated, murine amelogenin using a modified third generation genetic code expansion protocol. Protein Science. 2023;32(2):e4560. https://doi.org/10.1002/pro.4560

1469896x, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4560 by Oregon State University, Wiley Online Library on [05/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License