

# Amyloid-like amelogenin nanoribbons template mineralization via a low energy interface of ion binding sites

**Susrut Akkineni**

University of Washington

**Cheng Zhu**

University of Colorado

**Jiajun Chen**

University of Washington

**Miao Song**

Pacific Northwest National Laboratory

**Samuel Hoff**

Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, Colorado

**Johan Bonde**

Lund University

**Jinhui Tao**

Pacific Northwest National Laboratory

**Hendrik Heinz**

University of Colorado Boulder <https://orcid.org/0000-0002-6776-7404>

**Stefan habelitz**

University of California, San Francisco

**James De Yoreo** (✉ [james.deyoreo@pnnl.gov](mailto:james.deyoreo@pnnl.gov))

Pacific Northwest National Laboratory <https://orcid.org/0000-0002-9541-733X>

---

## Article

**Keywords:** amyloid-like amelogenin nanoribbons, mineralization, amorphous calcium phosphate nucleation, structure-function relationships

**Posted Date:** March 4th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-265667/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

# Abstract

Protein scaffolds direct the organization of amorphous precursors that transform into mineralized tissues, but the templating mechanism remains elusive. Inspired by a model of tooth enamel, wherein amyloid-like amelogenin nanoribbons guide apatite mineralization, we investigated the impact of nanoribbon structure and chemistry on amorphous calcium phosphate (ACP) nucleation. Using amelogenin sub-segments including an amyloid-like domain, nanoribbon conformation and function were determined by *in situ* atomic force microscopy and molecular dynamics simulations. All sequences substantially reduce nucleation barriers by creating low-energy interfaces, while phosphorylation dramatically enhances kinetic factors associated with ion binding. Furthermore, the predicted distribution of hydrophilic residues in the amyloid domain matches the structure of the multi-ion clusters comprising ACP. These findings provide crucial insights into structure-function relationships underlying amelogenin biomineralization and a generalizable system for synthesizing hybrid materials for various applications.

## Main

Hierarchical organic-inorganic hybrid structures, such as in tooth enamel and bone, are ubiquitous in biominerals, yet challenging to synthesize *in vitro*. In tooth enamel, mineralization is thought to begin with nucleation of amorphous calcium phosphate (ACP) in an amelogenin (Amel)-rich extracellular protein matrix and evolve to highly aligned bundles of apatite (AP) filaments with less than 2% residual protein<sup>1-3</sup>. Recent reports demonstrate that full-length Amel, an intrinsically disordered protein (IDP), can adopt an amyloid (cross  $\beta$ -sheet) quaternary structure and spontaneously self-assemble into nanoribbons (NR) 17 nm wide and a few nm in thickness<sup>4,5</sup>. *In vivo* studies revealed the presence of ribbon-like assemblies in the developing enamel matrix of various mammals, including humans<sup>4, 6-8</sup>, and these NR match the dimension and morphology of ribbons assembled from both recombinant human Amel (rH174)<sup>9</sup> and synthetic peptides<sup>4</sup>. *In vivo* observations suggest Amel NR are involved in mineralization of AP filaments which form during secretory stage of amelogenesis<sup>4</sup>, while recent *in vitro* experiments demonstrate that synthetic NR can indeed template growth of AP filaments starting with an amorphous precursor formed in the presence of acidic macromolecules<sup>5</sup>. However, the mechanism that leads to NR templating of ACP and its basis in NR and mineral structure remain unknown.

To determine the mechanism of mineral templating by human Amel NR, we performed *in situ* AFM observations of calcium phosphate nucleation on self-assembled NR of the full-length protein, as well as evolutionarily-conserved sub-segments hypothesized to be important for: 1) self-assembly; 2) stabilizing ACP (i.e., phosphorylated serine-16); and 3) binding with the AP surface (i.e., the C-terminus)<sup>4,11,12</sup>. To relate mechanism to structure we performed molecular dynamics (MD) simulations of the conformational states of NR formed from these peptide sub-segments.

Five sequences were designed. Recombinant full-length human Amel (rH174) and two peptide analogs, consisting of the domain that drives self-assembly (14P2) and 14P2 with truncated C-terminus appended

to it (14P2Cterm) shown in **Fig. 1, a to c**. In addition, phosphorylated versions of both peptides, p14P2 and p14P2Cterm, were also synthesized with phosphorylated serine sites highlighted in **Fig. 1, a to c**. The variable, histidine- and proline-rich central domain of Amel may also play a role in self-assembly or mineralization however its structure and function is unclear and was not investigated here.

All sequences assembled into NR on the surface of highly ordered pyrolytic graphite (HOPG) and remained stable in water (**Fig. 1, d to g**) for a pH range of 1.94 - 7.4 (Supplementary Fig. 1). NR dimensions for each sequence (Supplementary Table S1) were distinct, but films of the NR are similar in morphology, consisting of individual NR or islands with well aligned rows of NR. Further analysis of 14P2 (**Fig. 1d**) and other sequences (Supplementary Figs. 1b to e and 2) shows that, at high enough peptide or protein concentration, HOPG was completely covered, consisting of multiple layers of NR of 2 types, R1 and R2 (**Fig. 1i**) in contact with HOPG. R2 is the dominant morphology and can stack above one another (**Fig 1j**) whereas R1, rarely observed, forms a single highly crystalline layer (**Fig 1k**). R2 row widths and heights were roughly twice that of R1; e.g., for 14P2, the width and height of R1 rows are  $2.51 \pm 0.47$  nm and  $0.33 \pm 0.07$  nm, respectively, while those of R2 rows are  $5.81 \pm 0.60$  nm and  $0.65 \pm 0.04$  nm. In all cases, the NR were aligned with the three HOPG  $\langle 10\bar{1}0 \rangle$  directions (**Fig. 1f** and Supplementary Fig. 3).

Previous analysis of the  $\beta$ -sheet structure of amyloid ribbons by X-ray Diffraction (XRD) shows that dominant diffraction peaks lie at d-spacings of  $\sim 4.7$  Å and  $\sim 10$  Å<sup>13,14</sup>. The R2 NR prepared using our protocol exhibit similar diffraction patterns in bulk solution with and without HOPG, indicating that the structure is retained (Supplementary Fig. 4). To gain insight into the identity and distribution of exposed residues arising from the NR assembly, we performed MD simulations with the Interface-CHARMM force field with over 50 conformations for all peptides on HOPG, including virtual  $\pi$  electrons on graphite and aromatic amino acids which was shown to be critical to reproduce solvent and organic interfacial interactions (Supplementary Method 2)<sup>15-17</sup>.

The simulations predict the most favorable  $\beta$ -sheet conformation occurs in solution for both the basic 14P2 sequence (**Fig. 2a**) and the three modified versions (Supplementary Fig. 6). On HOPG, the standard  $\beta$ -sheet conformation is relatively less stable, with monomers beginning to spread after 7.5 ns (Supplementary Fig. 7). Further simulations with longer duration (10 ns) show that adopting a quasi- $\beta$ -sheet conformation with a slightly larger backbone gap in the central YINFSY domain ( $\sim 1$  nm vs  $\sim 0.5$  nm) and lower height ( $\sim 0.8$  nm vs.  $\sim 1.2$  nm) – nearly identical to R2 NR AFM topography – is most favorable, stabilizing the otherwise transient sheet (**Fig 2a** and Supplementary Fig. 8).

In both the transient and stabilized states, phenyl residues in the 14P2 domain are bound to the surface by  $\pi$ - $\pi$  interactions with graphite while adjacent, parallel ribbons formed hydrogen bonds between the N and C termini (Supplementary Fig. 8). In the stabilized conformation, hydrogen bonding is possible between -OH groups of Tyr<sub>5</sub> and Tyr<sub>10</sub> and the O=C- of the adjacent monomer backbone. Modification of 14P2 by phosphorylation (**Fig. 2b**) or appending the C-terminus (**Fig. 2c** and Supplementary Fig. 9) generated similar packing in the 14P2 domain. However, the DKTKREEVD domain extends into the solution, lacks the N and C terminus interaction with adjacent ribbons (Supplementary Fig. 8), and is

more disordered than the 14P2 domain (**Fig. 2c**). Further analysis reveals that nearly all hydrophilic side chains (in both transient and stabilized conformations), including phosphorylated serine, protrude into the solution with a periodicity that transverses the long axis of the nanoribbon (thick lines in **Fig. 2a, b** and **c**).

To quantify the impact of this structural motif on calcium phosphate nucleation rates, we used *in situ* AFM to investigate nucleation using substrates with the dominant R2 NR, identical to those in **Fig. 1**, at pH 7.4 and 25 °C for a range of supersaturations  $\sigma$  (Supplementary Tables 2 and 4) over timescales for which nucleation was absent both in solution and on bare HOPG (Supplementary Fig. 10). The value of  $\sigma$  was determined in two ways (Supplementary Table S4): from the value of the equilibrium constant  $K_{sp}$  reported previously at a different ionic strength<sup>18</sup> and using the ionic activity product at which the post-nucleation growth rate extrapolated to zero (Supplementary Method 8).

Time-lapse images (**Fig. 3, a to e** and Supplementary Videos 1 to 5) revealed that ACP ( $(Ca_2(HPO_4)_3)^{2-}$ ) was the first phase to form at all values of  $\sigma$  explored here for all five sequences, as validated by *in situ* AFM and electron microscopy (Supplementary Figs. 11 to 13) and observed previously for nucleation on a number of proteins<sup>18-20</sup>. The ACP particles grew in size (Supplementary Table S3) before transforming to fiber- or plate-shaped mineral (**Fig. 3, a to e panel 5** and Supplementary Fig. 13).

Nucleation rates and total mineral coverage were highest on the phosphorylated peptides, p14P2 and p14P2Cterm (**Fig. 3 f** and Supplementary Fig. 14, Supplementary Tables 5 and 6), which had particles concentrated along the length and direction of the NR (Supplementary Video 6). In contrast, ACP growth rates on non-phosphorylated peptides were higher than on the phosphorylated versions (**Fig. 3g** and Supplementary Fig. 14, Supplementary Table S3). Furthermore, appending the C-terminus to 14P2 and p14P2 sequences resulted in an increase of 1.3 to 2 times in nucleation rates and a factor of 1.5 to 2.3 in (vertical) growth rates. Comparison to previous results using conformationally dissimilar Amel nanospheres and nanospheres co-assembled with highly charged enamelin (Enam) (**Fig. 3h, Amel sph: Enam, 50:0 and 50:1**)<sup>19</sup> shows that all NR sequences drive higher nucleation rates than do nanospheres, and the phosphorylated NR outperform the mixed Amel nanosphere-Enam system, even at lower supersaturations and for the optimal Amel:Enam ratio. In addition, phosphorylated NR induced ACP nucleation rates 1.9 to 2.8 times larger than observed on collagen (**Fig. 3h, Col.  $\sigma_{ACP} = 0.128$** )<sup>18</sup>. These findings show that both sequence *and* conformation strongly impact template activity and that phosphorylated NR are the most potent ACP nucleators investigated to date.

To determine the mechanism and underlying energetic factors through which Amel NR drive ACP nucleation, the data on nucleation rates vs  $\sigma$  were analyzed using classical nucleation theory (CNT), which has been used previously to analyze heterogeneous nucleation on organic templates<sup>18,19,21-23</sup> and has been shown to effectively describe ACP nucleation kinetics<sup>18</sup>. CNT predicts that the heterogeneous nucleation rate ( $J_0$ ) varies exponentially with the effective interfacial energy ( $\alpha_{ACP}$ ) and  $\sigma_{ACP}$  according to:

$$\ln(J_o) = \ln(A) - \frac{B}{\sigma_{ACP}^2} \quad (1)$$

$$B = \frac{8\pi\omega^2\alpha_{ACP}^3}{3(kT)^3} \quad (2)$$

where  $A$  is the kinetic pre-factor, which depends on the density of possible nucleation sites (24), on attachment rates, and barriers to ion desolvation<sup>25</sup>,  $\omega$  is volume of the ACP growth unit,  $k$  is Boltzmann's constant and  $T$  is absolute temperature. (See Supplementary Method 10).

Fitting the data for all values of  $\sigma_{ACP}$  (Fig. 3i and Supplementary Fig. 15, Supplementary Tables 7 and 8) results in values of  $\alpha_{ACP}$  for all NR sequences that are remarkably small, ranging from 1.4 - 2.4 mJ m<sup>-2</sup> or 12 - 20 mJ m<sup>-2</sup> (depending on the method used to determine  $K_{sp}$ ), when compared to either nucleation on collagen<sup>18</sup> (40 mJ m<sup>-2</sup>) or in bulk solution<sup>18</sup> (~150 mJ m<sup>-2</sup>). Hence, we infer that nucleation is highly favorable on Amel NR due to the low interfacial energy. The data also show that both addition of the C-terminus to 14P2 and phosphorylation of the analogs have minor effects on  $\alpha_{ACP}$ , with the induced change ranging from a factor of 0.9 to 1.5 (Fig. 3j and Supplementary Tables 7 and 8). In contrast, phosphorylation results in a 15 to 19-fold increase in the kinetic pre-factor  $A$ . Comparison of the value of  $A$  with the net charge of the peptide reveals a strong linear trend, likely reflecting rates of ion binding or the lifetime of bound ions.

To understand the structural reason for why phosphorylated 14P2 NR so dramatically enhances nucleation kinetics, we compared the geometric arrangement of the NR's charged side chains with the dimers of the Ca(HPO<sub>4</sub>)<sub>3</sub><sup>4-</sup> multi-ion complexes shown previously to be the incipient species leading to formation of ACP through binding of an additional Ca<sup>2+</sup> ion, as well as to octacalcium phosphate (OCP) and AP<sup>18</sup> through further ion binding reactions. Superposition of Ca<sup>2+</sup> ions in the dimeric complex onto the p14P2 NR conformation (stabilized or transient sheet) shows an excellent match with the spacing (~1 nm) of the periodic hydrophilic residues that span the long axis of the NR (Fig 4a). Thus, binding of Ca<sup>2+</sup> ions to the periodic phosphorylation sites may provide a means to stabilize the calcium triphosphate dimers and promote formation of ACP beyond that afforded alone by a template of low interfacial energy.

The above experimental findings show that the amyloid structure of Amel NR, especially the evolutionarily conserved 14P2 self-assembly domain<sup>26</sup>, drastically lowers the thermodynamic barrier for nucleation of ACP, while the simulations predict that Amel NR present a periodic template of hydrophilic residues for mineral formation – without involvement of non-Amel proteins (Fig. 4b). Moreover, phosphorylation of the NRs dramatically enhances the kinetic pre-factor, likely through enhanced Ca<sup>2+</sup> ion binding at charged residues. Addition of the Amel C-terminus has little impact on these two factors, suggesting that amyloid-like Amel NR can retain its structure and function even after C-terminal cleavage. Whether the C-terminus then independently acts to inhibit HAP (100) growth as previously proposed<sup>27</sup>, or acts as a highly charged macromolecule to produce the polymer induced liquid precursor (PILP) used in previous NR mineralization experiments to produce aligned AP filaments<sup>8</sup>, remains to be explored.

The results also provide evidence that one region of a matrix IDP can fold to drive self-assembly into a template that stabilizes a mineral phase while leaving another region unfolded to interact with non-matrix proteins, ions or crystals<sup>28,29</sup>. The use of such sequence-specific interfacial behavior towards crystal nucleation can, in principle, be expanded beyond proteins to block copolymers, peptoids and other amphiphilic polymers in which different domains can control distinct aspects of nucleation through reduction of mineral-binding energies by assembly into an ordered template and enhancement of ion-binding at exposed charged residues<sup>30-32</sup>.

The link between the self-assembled molecular structure of Amel NR templates and mineral nucleation points towards a range of translational opportunities. The ability to readily assemble peptides with control over orientation on crystalline surfaces like graphene<sup>33</sup>, mica<sup>34</sup> or MoS<sub>2</sub><sup>16</sup>, or to array proteins on surfaces patterned at the nanoscale using block copolymers films<sup>35</sup> or nanolithography<sup>37</sup>, provides a potential path to organize Amel NR for dental tissue engineering. Alternatively, the periodic organization of nucleation sites on the NRs suggests a means for creating patterns of quantum dots of other materials by tailoring the site chemistry. More broadly, achieving predictive hierarchical growth of inorganic crystalline arrays directed by organic scaffold, inspired by nature, is a long sought-after vision of bioinspired material science. The robust hierarchical design of amyloid NRs arrayed with modular functional groups around a self-assembling domain, provides a general scaffold design upon which that vision may be realized.

## Methods

### Synthesis of rH174 and peptide analogs

Recombinant human amelogenin (rH174) was expressed in *Escherichia coli* BL21 (DE3) and purified according to a previously reported protocol<sup>37</sup>. In short, soluble rH174 was extracted from the cells using an acid/heat treatment, which was followed by extensive dialysis against 0.05% acetic acid, and subsequent lyophilization. The purity of the protein was above 95% as assessed by HPLC and SDS-PAGE. This rH174 lacks the first amino acid residue (Met1) and phosphorylation at Ser16 compared to native human amelogenin. The peptides were synthesized, purified with HPLC (95% purity), and lyophilized at Elim Biopharma Inc (Hayward, CA).

### Mineral-free self-assembly

1 mg of the lyophilized protein or peptide was dissolved in 1 ml of 10 mM 99.999% HCl (Sigma-Aldrich), pH 1.94, followed by 30 minutes of sonication in a Branson ultrasonic bath (M5800H, 40 kHz frequency) to disperse the lyophilized protein flakes, and then vortexed for 30 seconds to obtain a stock solution concentration of 1 mg/ml. Concentration of 14P2 was 626.2  $\mu$ M, p14P2 was 596  $\mu$ M, 14P2Cterm was 370.75  $\mu$ M, p14P2Cterm was 360  $\mu$ M and rH174 was 50.08  $\mu$ M. For lower concentrations, dilution was performed immediately with 10 mM HCl. Aliquoted solutions were kept at room temperature (25 °C) during AFM experiments or stored at 4 °C, viable for at least 30 days.

### AFM substrates for mineralization

Highly ordered pyrolytic graphite (HOPG) ZYB quality was purchased from Ted Pella, mounted on metal AFM stubs (Ted Pella) using epoxy and the exposed side was cleaved with scotch tape immediately before use. The protein or peptide solutions were used after 48 hrs of incubation at room temperature. 60  $\mu\text{l}$  of the protein or peptide solution (concentrations: 14P2: 31.31  $\mu\text{M}$ , p14P2: 59.67  $\mu\text{M}$ , 14P2Cterm: 37.08  $\mu\text{M}$ , p14P2Cterm: 36  $\mu\text{M}$ , and rH174: 5  $\mu\text{M}$ ) was drop-casted on the surface. This substrate was incubated at 37 °C in a sealed chamber (RH ~ 100%) for 24 hrs. After incubation, the drop-casted solution was exchanged with 1 mM HCl (pH 3.1) for three times to remove the unbound protein. Then, the acidic buffer was exchanged with deionized H<sub>2</sub>O and 40  $\mu\text{l}$  of H<sub>2</sub>O was left on the surface in order to prevent drying. The substrate was stored in a sealed humid chamber (RH ~ 100%) at room temperature until further use. Concentration of proteins was optimized for full surface coverage over several experiments. For peptides, concentration above ~ 0.3  $\mu\text{M}$  was sufficient for the fully covered monolayer and higher concentrations resulted in additional layers. The number of layers did not affect nucleation rates during nucleation experiments (data not shown) as long as the surface was fully covered. For rH174, low concentrations were required to prevent the formation of large globular aggregates in high concentrations.

### Molecular dynamics simulations

Materials Studio was utilized to build all-atom models of the peptides and HOPG surfaces. Peptide models were built based on the peptide sequences GHPGYINFSYEVLT (14P2), GHPGYINF(p)SYEVLT (p14P2), GHPGYINFSYEVLTDKTKREEVD (14P2Cterm), and GHPGYINF(p)SYEVLTDKTKREEVD (p14P2Cterm). See Supplementary Method 2 for detailed methods and discussion.

### In situ AFM experiments

#### 1. For monomer resolution AFM

Silicon nitride tip cantilevers (Bruker, OTR4, spring constant k: 0.08 N/m or OTR8, spring constant k: 0.15 N/m) were plasma treated for 1 min before use to make them hydrophilic. Substrates were placed on the AFM stage (Cypher VRS, Asylum Research) and surface was imaged in liquid at room temperature (25°C). Freshly made peptide solutions were used. Images of 14P2 R1 nanoribbons (Fig. 1i and k) were obtained by drop casting 6.26  $\mu\text{M}$  onto freshly cleaved HOPG surface and immediately imaged using tapping mode. Similar process was used for images of R2 nanoribbons with 31.31  $\mu\text{M}$  (Fig. 1j) by using contact mode.

#### 2. For nucleation and growth with supersaturated solutions

Silicon nitride cantilevers with Si tip (Applied Nanostructures HYDRA4V-100NG, spring constant k: 0.088 N/m or Bruker SNL-10 C, spring constant k: 0.24 N/m) were treated with UV/ozone for 15 minutes before use to remove residual organic contaminants and improve hydrophilicity. Substrates were placed on the AFM stage (MultiMode 8, Bruker), operated in tapping mode and surface was imaged in liquid at room temperature (25°C). Protocol for mineralization experiments and analysis is identical to procedures for

heterogeneous nucleation, growth and phase transformation with collagen and amelogenin functionalized mica in previous publications<sup>18,19</sup>. Three sets of calcium and phosphate solutions, 2 mM CaCl<sub>2</sub> and 19 mM KH<sub>2</sub>PO<sub>4</sub> ( $\sigma_{AP} = 3.37$ ), 1.75 mM CaCl<sub>2</sub> and 16.95 mM KH<sub>2</sub>PO<sub>4</sub> ( $\sigma_{AP} = 3.28$ ), and 1.5 mM CaCl<sub>2</sub> and 14.9 mM KH<sub>2</sub>PO<sub>4</sub> ( $\sigma_{AP} = 3.19$ ) before mixing, were calculated using Visual MINTEQ while maintaining log (activity of Ca<sup>2+</sup>/activity of PO<sub>4</sub><sup>3-</sup>) of 3.87 and solutions were independently prepared. pH of solutions was adjusted to 7.4 using concentrated KOH or HCl, then filtered three times with a cellulose acetate filter (pore size of 0.22  $\mu$ m) before immediate use.

The filtered CaCl<sub>2</sub> and KH<sub>2</sub>PO solutions were pumped at 25  $\mu$ l/min for  $\sigma_{AP} = 3.37$ , 31.25  $\mu$ l/min for  $\sigma_{AP} = 3.28$  and 37.25  $\mu$ l/min for  $\sigma_{AP} = 3.19$ , using a syringe pump (Harvard Apparatus) and inert tubing, and finally combined at the inlet of the AFM liquid cell using a custom T-junction with minimal dead volume and residence time. The flow rates were experimentally optimized to avoid diffusion-limited nucleation and solutions were buffered by high phosphate concentration for at least 3 hours after mixing. Due to continuous flow of buffered phosphate ions and high flow rates, the system is at constant composition during nucleation stage as observed in previous reports and no bulk solution-formed precipitates deposited<sup>18,19</sup>. The scan size, rate and sample lines/frame were optimized with several experiments for each sequence and supersaturation to optimize the imaging time and image quality.

#### AFM image processing

Raw height images were processed using a polynomial flatten function and baselines corrected on Gwyddion SPM data analysis software. In addition, a conservative gaussian smoothing filter was applied to Fig. 1d to h and Supplementary Videos 1 to 6 to remove pixelation from magnification. Particle counting and surface coverage analysis was performed with ImageJ (NIH).

#### Nucleation rate and density analysis

Nucleation images were analyzed using ImageJ (NIH). Strict classification rules were placed for definition of a nuclei, i.e., particles that appear in a frame and stay on to grow in the subsequent frames. However, in certain frames, nuclei disappeared or translated to another position after growth over the course of the experiment from the original locations. These nuclei were also counted and carefully tracked, frame by frame. The nuclei number densities in each frame were calculated by dividing number with frame sizes and plotted against time captured at end of the scan. The relationship of nuclei number density vs elapsed time was then linearly fit to get the slope, which is defined as nucleation rate for each supersaturation on a specific peptide or protein sequence. (Supplementary Table 5).

#### Net charge of monomers at pH 7.4

Net charge for each sequence was calculated (Supplementary Tables 7 and 8) using the bioinformatics toolbox, Prot pi, available at <https://www.protpi.ch/Calculator/ProteinTool>. Net charge relative to 14P2 reflects Net Charge of sequence/Net charge of 14P2.

## Statistics

## Growth velocity, Fig. 3g

Average particle height was measured for  $n$  particles that had minimum number of neighboring particles. Standard deviation and standard error were calculated. Error bar in Fig. 3g is standard error (range).

$\sigma_{AP} = 3.19$ ,  $\sigma_{ACP} = 0.04$ : 14P2  $n = 5$ ; p14P2  $n = 6$ ; 14P2Cterm  $n = 4$ ; p14P2Cterm  $n = 4$ ; rH174  $n = 9$

$\sigma_{AP} = 3.28$ ,  $\sigma_{ACP} = 0.138$ : 14P2  $n = 10$ ; p14P2  $n = 8$ ; 14P2Cterm  $n = 7$ ; p14P2Cterm  $n = 6$ ; rH174  $n = 5$

$\sigma_{AP} = 3.37$ ,  $\sigma_{ACP} = 0.221$ : 14P2  $n = 7$ ; p14P2  $n = 8$ ; 14P2Cterm  $n = 7$ ; p14P2Cterm  $n = 6$ ; rH174  $n = 7$

## Nucleation rates and surface energy: Fig. 3, f and i and Supplementary 15a

Nucleation was repeated at least 5 times on each sequence for each supersaturation. Data sets with modal slope values, negligible instrumental artifacts and minimal variations in substrate topography were selected to calculate averages, and error bars represent standard deviation.

$\sigma_{AP} = 3.19$ ,  $\sigma_{ACP} = 0.04$ : 14P2  $n = 2$ ; p14P2  $n = 2$ ; 14P2Cterm  $n = 2$ ; p14P2Cterm  $n = 2$ ; rH174  $n = 2$

$\sigma_{AP} = 3.28$ ,  $\sigma_{ACP} = 0.138$ : 14P2  $n = 3$ ; p14P2  $n = 3$ ; 14P2Cterm  $n = 2$ ; p14P2Cterm  $n = 3$ ; rH174  $n = 2$

$\sigma_{AP} = 3.37$ ,  $\sigma_{ACP} = 0.221$ : 14P2  $n = 4$ ; p14P2  $n = 2$ ; 14P2Cterm  $n = 4$ ; p14P2Cterm  $n = 2$ ; rH174  $n = 2$

## Relative interfacial energy and kinetic pre-factor: Fig. 3, j and k and Supplementary 15, b and c

Error bars represent standard error (range). Error for kinetic pre-factor  $A$  was calculated from error of the intercept  $\ln(A)$  in Fig. 3i and Supplementary Fig. 15a, using the expression: Error of  $A = (e^{\text{Intercept}} \times \text{Error of intercept from fitting})$ . Errors of the ratios for each modified sequence reflects the error range of the modified sequence and that of 14P2.

## **Declarations**

### **Acknowledgements**

Experiments were performed at the Pacific Northwest National Laboratory operated by Battelle for the U.S. Department of Energy (DOE) under contract DE-AC05-76RL01830, Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program, and Beamline 8.3.1 of the Advanced Light Source, a U.S. DOE Office of Science User Facility under Contract No. DE-AC02-05CH11231, supported in part by the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01. Funding: Supported by NIH/NIDCR Grant R01-DE025709. Simulations were performed using NSF-OAC-1931587, NSF-CMMI-1940335, the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357, and the Summit supercomputer, a joint effort of the University of Colorado Boulder and Colorado State University, which is supported by the

National Science Foundation (ACI-1532235 and ACI-1532236). Authors also acknowledge support from Dr. Paul Ashby and helpful discussions with Dr. Wu Li, Dr. Elias Nakouzi, Dr. Benjamin Legg and Dr. George Meigs.

### Author contributions

Conceptualization: SA, CZ, JT, HH, SH, JJD; Methodology: SA, JT, CZ, SEH, SH, HH, JJD; Analysis: SA, JT, CZ, JJD; Investigation: SA, CZ, JT, JC, MS, SEH; Writing-Original Draft: SA, CZ, JT, and JJD with feedback from all authors; Supervision: JT, HH, SH and JJD.

### Competing interests

Authors declare no competing interests.

### Data and materials availability

All data are available by contacting the corresponding authors

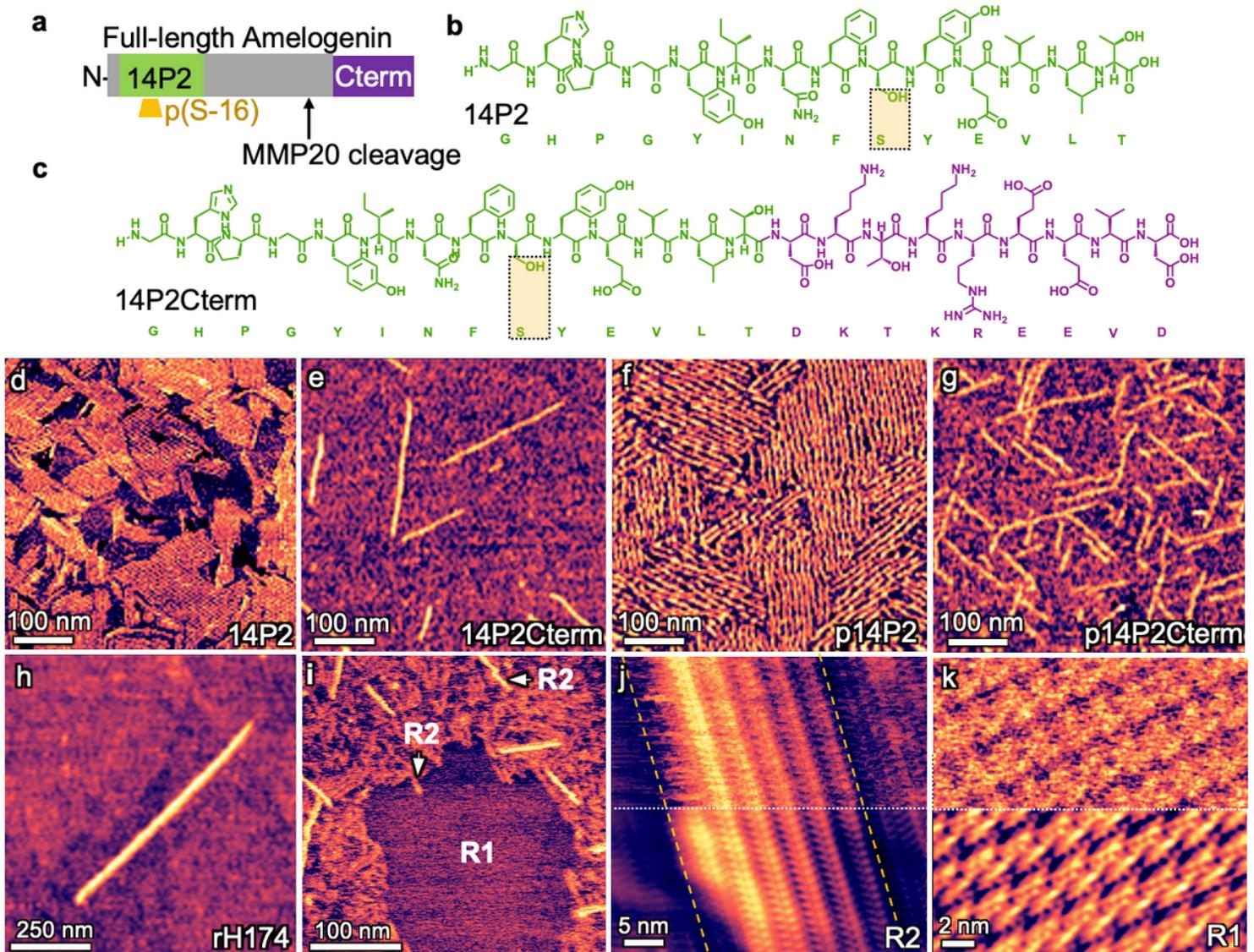
## References

1. Smith, C. E., Hu, Y., Hu, J. C.-C. & Simmer, J. P. Ultrastructure of early amelogenesis in wild-type, *Amelx*<sup>-/-</sup>, and *Enam*<sup>-/-</sup> mice: enamel ribbon initiation on dentin mineral and ribbon orientation by ameloblasts. *Mol. Genet. genomic Med.* **4**, 662–683 (2016).
2. Beniash, E., Metzler, R. A., Lam, R. S. K. & Gilbert, P. U. P. A. Transient amorphous calcium phosphate in forming enamel. *J. Struct. Biol.* **166**, 133–143 (2009).
3. Pandya, M. *et al.* Posttranslational amelogenin processing and changes in matrix assembly during enamel development. *Front. Physiol.* **8**, (2017).
4. Carneiro, K. M. M. *et al.* Amyloid-like ribbons of amelogenins in enamel mineralization. 1–11 (2016).
5. Martinez-Avila, O. *et al.* Self-Assembly of Filamentous Amelogenin Requires Calcium and Phosphate: From Dimers via Nanoribbons to Fibrils. *Biomacromolecules* **13**, 3494–3502 (2012).
6. Pautard, F. G. E. An X-ray diffraction pattern from human enamel matrix. *Arch. Oral Biol.* **3**, 217-IN13 (1961).
7. Jodaikin, A., Traub, W. & Weiner, S. Protein conformation in rat tooth enamel. *Arch. Oral Biol.* **31**, 685–689 (1986).
8. Bai, Y. *et al.* Protein nanoribbons template enamel mineralization. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 19201–19208 (2020).
9. Sanii, B., Martinez-Avila, O., Simpliciano, C., Zuckermann, R. N. & Habelitz, S. Matching 4.7-Å XRD spacing in amelogenin nanoribbons and enamel matrix. *J. Dent. Res.* **93**, 918–22 (2014).
10. Zhang, J., Wang, J., Ma, C. & Lu, J. Hydroxyapatite Formation Coexists with Amyloid-like Self-Assembly of Human Amelogenin. *Int. J. Mol. Sci.* **21**, 2946 (2020).

11. Shin, N. Y. *et al.* Amelogenin phosphorylation regulates tooth enamel formation by stabilizing a transient amorphous mineral precursor. *J. Biol. Chem.* **295**, 1943–1959 (2020).
12. Friddle, R. W. *et al.* Single-molecule determination of the face-specific adsorption of Amelogenin's C-terminus on hydroxyapatite. *Angew. Chem. Int. Ed. Engl.* **50**, 7541–5 (2011).
13. Engelberth, S. A. *et al.* Progression of Self-Assembly of Amelogenin Protein Supramolecular Structures in Simulated Enamel Fluid. *Biomacromolecules* **19**, 3917–3924 (2018).
14. Ma, C.-W., Zhang, J., Dong, X.-Q. & Lu, J.-X. Amyloid structure of high-order assembly of Leucine-rich amelogenin revealed by solid-state NMR. *J. Struct. Biol.* **206**, 29–35 (2019).
15. Heinz, H., Lin, T.-J., Kishore Mishra, R. & Emami, F. S. Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field. *Langmuir* **29**, 1754–1765 (2013).
16. Chen, J. *et al.* Building two-dimensional materials one row at a time: Avoiding the nucleation barrier. *Science (80-.)*. **362**, 1135–1139 (2018).
17. Pramanik, C., Gissinger, J. R., Kumar, S. & Heinz, H. Carbon Nanotube Dispersion in Solvents and Polymer Solutions: Mechanisms, Assembly, and Preferences. *ACS Nano* **11**, 12805–12816 (2017).
18. Habraken, W. J. E. M. *et al.* Ion-association complexes unite classical and non-classical theories for the biomimetic nucleation of calcium phosphate. *Nat. Commun.* **4**, 1507 (2013).
19. Tao, J. *et al.* Control of Calcium Phosphate Nucleation and Transformation through Interactions of Enamelin and Amelogenin Exhibits the 'Goldilocks Effect'. (2018).
20. Chien, Y.-C. *et al.* Using biomimetic polymers in place of noncollagenous proteins to achieve functional remineralization of dentin tissues. *ACS Biomater. Sci. Eng.* **3**, 3469–3479 (2017).
21. Hu, Q. *et al.* The thermodynamics of calcite nucleation at organic interfaces: Classical vs. non-classical pathways †. *Faraday Discuss* **159**, 509–523 (2012).
22. Giuffre, A. J., Hamm, L. M., Han, N., De Yoreo, J. J. & Dove, P. M. Polysaccharide chemistry regulates kinetics of calcite nucleation through competition of interfacial energies. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9261–6 (2013).
23. Hamm, L. M. *et al.* Reconciling disparate views of template-directed nucleation through measurement of calcite nucleation kinetics and binding energies. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1304–9 (2014).
24. Kashchiev, D. & van Rosmalen, G. M. Review: Nucleation in solutions revisited. *Cryst. Res. Technol.* **38**, 555–574 (2003).
25. De Yoreo, J. J. & Vekilov, P. G. Principles of Crystal Nucleation and Growth. *Rev. Mineral. Geochemistry* **54**, 57–93 (2003).
26. Wald, T. *et al.* Intrinsically disordered proteins drive enamel formation via an evolutionarily conserved self-assembly motif. doi:10.1073/pnas.1615334114
27. Iline-Vul, T. *et al.* Osteopontin regulates biomimetic calcium phosphate crystallization from disordered mineral layers covering apatite crystallites. *Sci. Rep.* **10**, 15722 (2020).

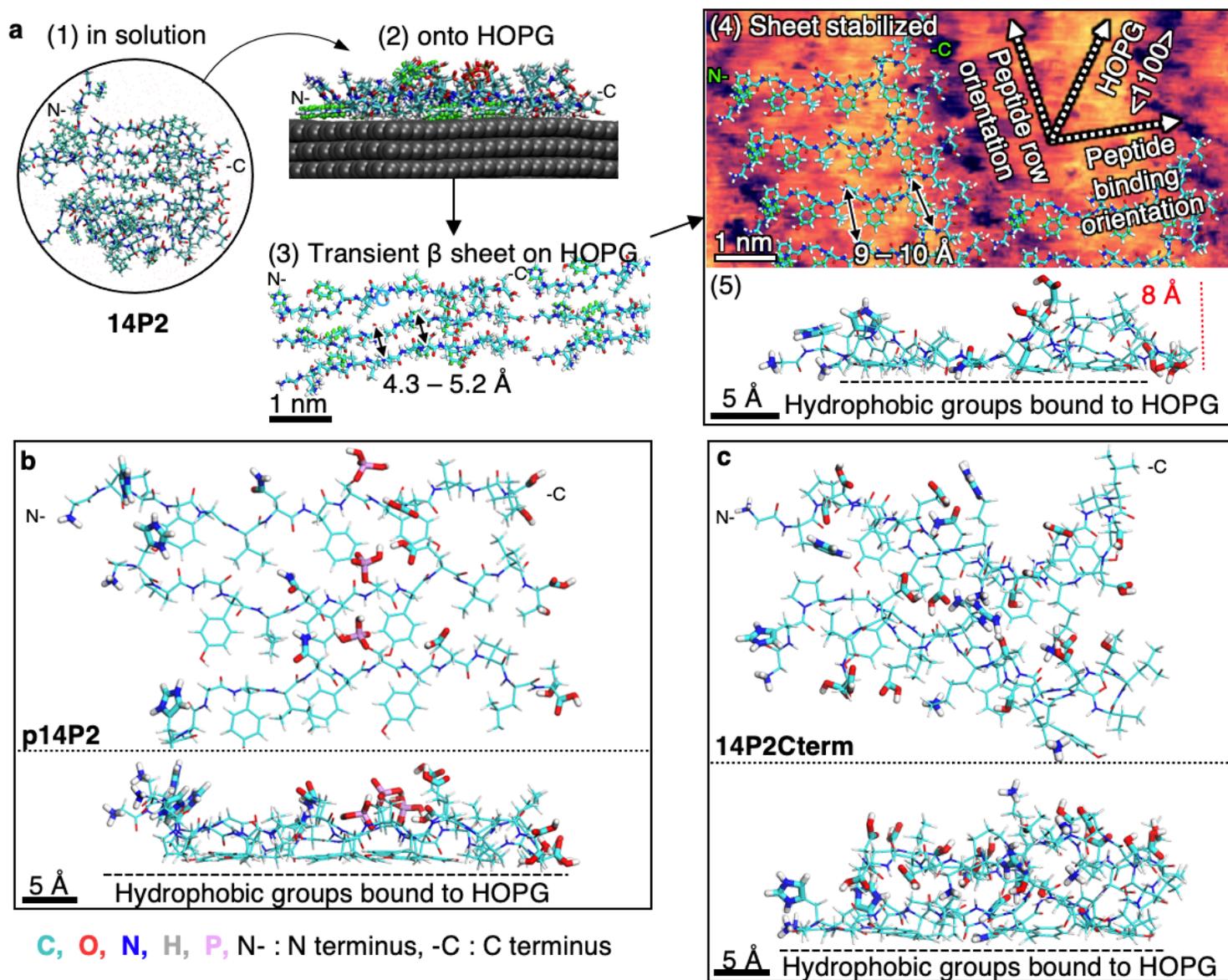
28. Boskey, A. L. & Villarreal-Ramirez, E. Intrinsically disordered proteins and biomineralization. *Matrix Biol.* **52–54**, 43–59 (2016).
29. Rao, A. *et al.* Stabilization of Mineral Precursors by Intrinsically Disordered Proteins. *Adv. Funct. Mater.* **28**, 1802063 (2018).
30. Oleske, K. W. *et al.* Nanopatterning of Crystalline Transition Metal Oxides by Surface Templated Nucleation on Block Copolymer Mesostructures. *Cryst. Growth Des.* **17**, 5775–5782 (2017).
31. Kim, Y.-Y. *et al.* Hydroxyl-rich macromolecules enable the bio-inspired synthesis of single crystal nanocomposites. *Nat. Commun.* **10**, 5682 (2019).
32. Yan, F. *et al.* Controlled synthesis of highly-branched plasmonic gold nanoparticles through peptoid engineering. *Nat. Commun.* **9**, 2327 (2018).
33. Sarikaya, M., Tamerler, C., Jen, A. K.-Y., Schulten, K. & Baneyx, F. Molecular biomimetics: nanotechnology through biology. *Nat. Mater.* **2**, 577–585 (2003).
34. Pyles, H., Zhang, S., De Yoreo, J. J. & Baker, D. Controlling protein assembly on inorganic crystals through designed protein interfaces. *Nature* **571**, 251–256 (2019).
35. Stel, B. *et al.* Contrasting Chemistry of Block Copolymer Films Controls the Dynamics of Protein Self-Assembly at the Nanoscale. *ACS Nano* **13**, 4018–4027 (2019).
36. Lee, K.-B., Park, S.-J., Mirkin, C. A., Smith, J. C. & Mrksich, M. Protein Nanoarrays Generated By Dip-Pen Nanolithography. *Science (80-)*. **295**, 1702–1705 (2002).
37. Svensson Bonde, J. & Bulow, L. One-Step Purification of Recombinant Human Amelogenin and Use of Amelogenin as a Fusion Partner. *PLoS One* **7**, e33269 (2012).

## Figures



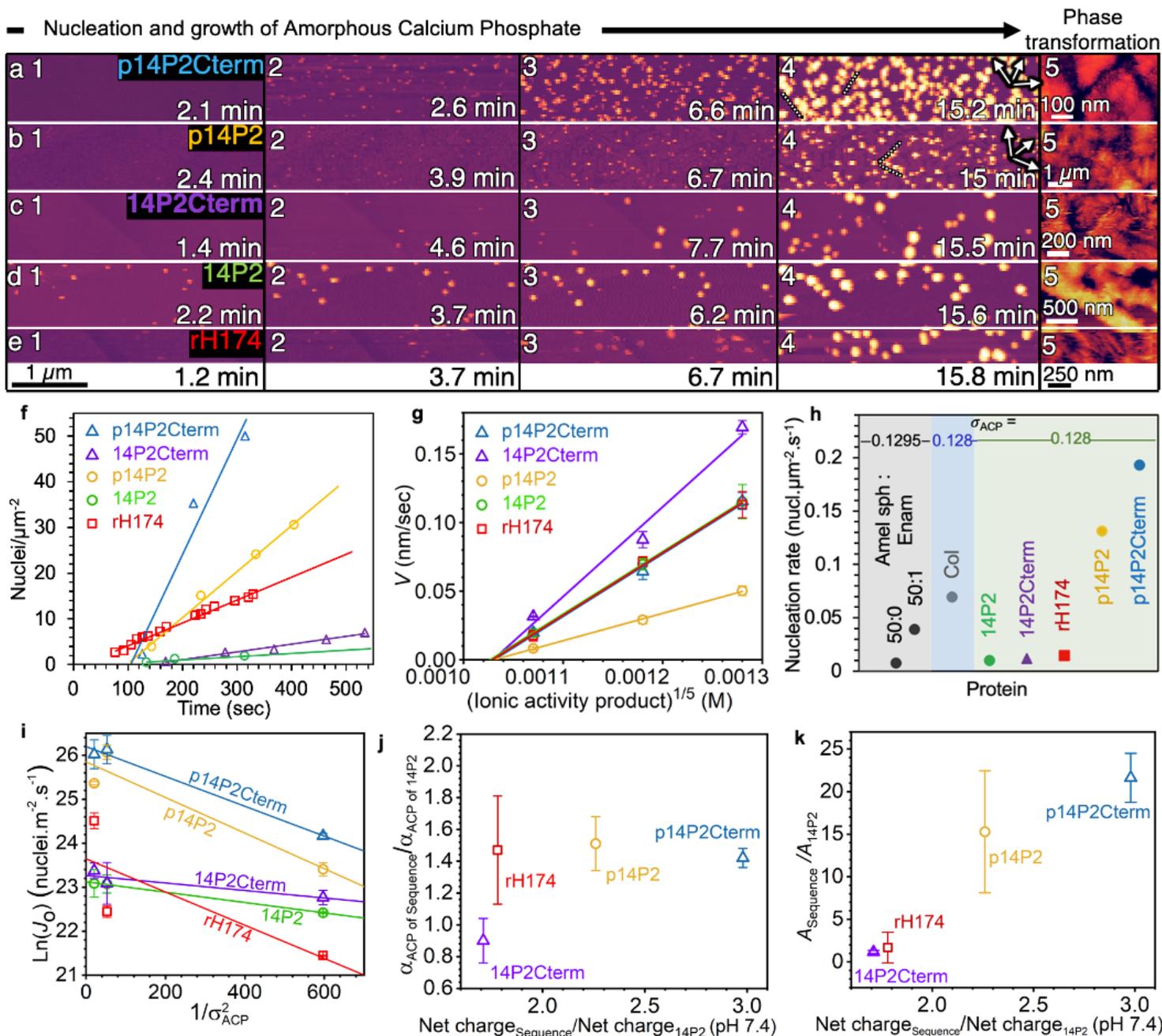
**Figure 1**

Sequences and assemblies of full-length Amel and peptide analogs on HOPG. (a) Location of functional domains in human Amel, (b) self-assembling 14P2 domain, (c) 14P2Cterm (14P2 appended with truncated C-terminus, which is important for AP binding). (d-h) In situ AFM images of ordered nanoribbons assembled on HOPG and characterized in pure water: (d) 14P2; (e) 14P2Cterm; (f) and (g) phosphorylated versions p14P2 and p14P2Cterm, respectively; (h) recombinant full-length rH174 Amel. (i) In situ AFM on HOPG in 6.26  $\mu\text{M}$  14P2 solution reveals 100% coverage with multiple layers: two types, R1 and R2. (j) Structure of R2 in 31.3  $\mu\text{M}$  14P2 solution; dashed lines delineate the boundaries of a second layer over the first. (k) Structure of R1 in indicated region of (i). (j) and (k) top: original image; bottom: processed Fast-Fourier transform filter.



**Figure 2**

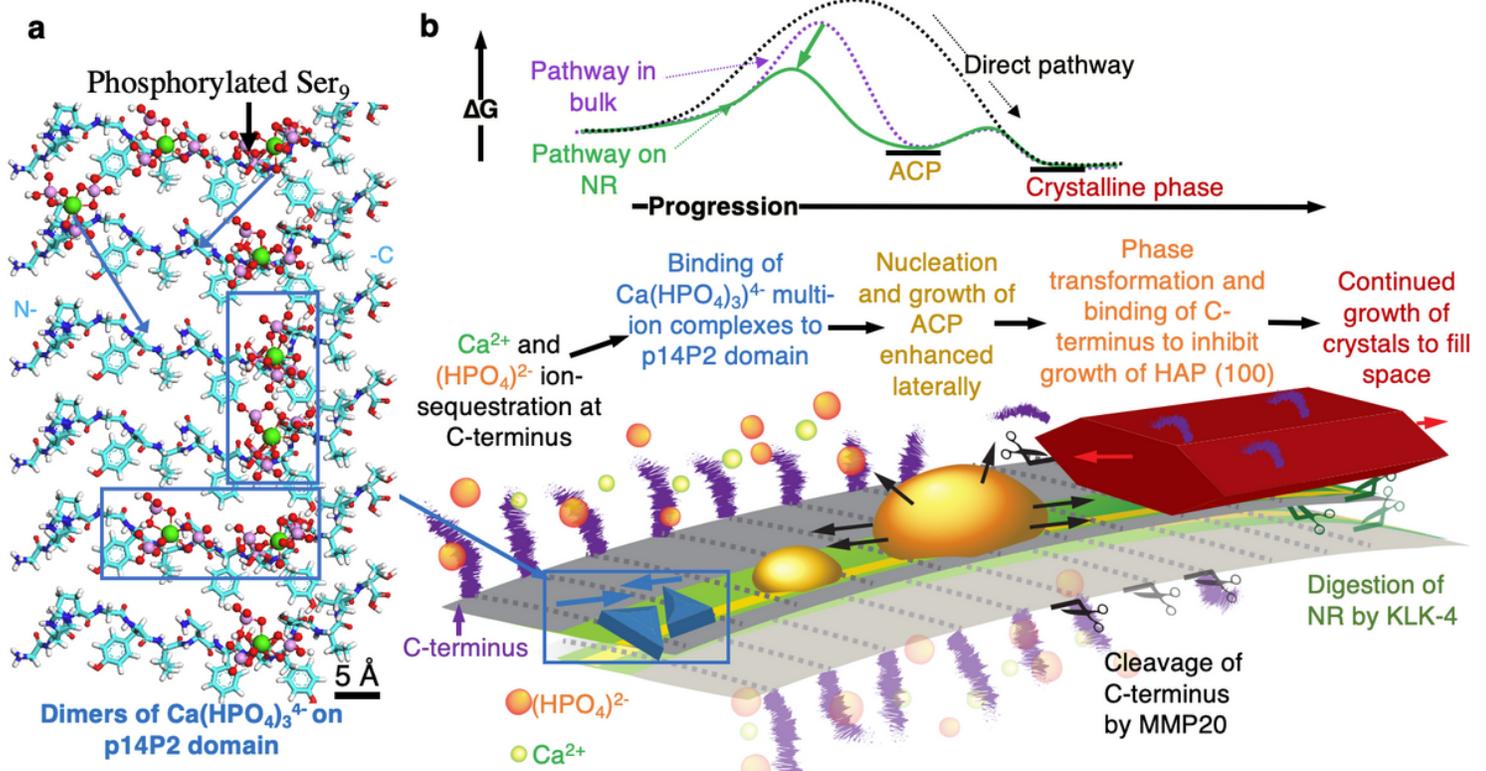
Lowest energy conformation of peptides in solution and on HOPG (0001) from all-atom MD simulations. (a) Schematic of procedure for simulations: 1. Snapshot of lowest energy conformations for 14P2 peptides generated for  $\beta$ -sheet structure in solution; 2. Side-view of transient  $\beta$ -sheet on HOPG (height  $\sim 1.2$  nm); 3. Top-down view of transient  $\beta$ -sheet on HOPG after 7.5 ns; 4. Stabilized sheet on HOPG simulated for 10 ns with larger gap between monomer backbones overlaid on AFM image of R2; and 5. height reduces (side view) due to  $\pi$ - $\pi$  interactions with HOPG. Snapshots of top and side view for stabilized sheet on HOPG after 8 ns for (b) p14P2 and (c) 14P2Cterm. HOPG is not shown for clarity. All peptide sequences have well-ordered 14P2 domains and hydrophilic sites (thicker lines), including phosphorylated serine, that face the solution.



**Figure 3**

Kinetics and energetics of ACP nucleation on peptide and protein nanoribbons. (a-e) Time lapse of AFM images using constant supersaturation,  $\sigma_{AP} = 3.37$ ,  $\sigma_{ACP} = 0.221$ , at 25° C and pH 7.4, with  $t = 0$  min defined as the time when solution is introduced into the flow cell. Dotted lines indicate ACP aligned with direction of NR (arrows). (f) Plot of nuclei number density over time measured for all sequences at  $\sigma_{ACP} = 0.22$ . Error bar is smaller than data points. (g) Average growth rate of particle height ( $V$ ) at  $\sigma_{ACP} = 0.04$ , 0.138 and 0.221. (h) Comparison of nucleation rates for different proteins: amelogenin nanospheres co-assembled without (50:0) and with enamelin (50:1) at  $\sigma_{ACP} = 0.1295$ , collagen at  $\sigma_{ACP} = 0.128$  and amelogenin nanoribbons at  $\sigma_{ACP} = 0.128$  (from fits to data in (i)). (i) Linear fits to  $\ln(J_0)$  (nuclei m<sup>-2</sup> s<sup>-1</sup>) measured at different ACP supersaturations ( $1/\sigma_{ACP}^2$ ) according to Equation 1. (j) Ratio of interfacial energy ( $\alpha_{ACP}$ ) and (k) ratio of kinetic pre-factor ( $A$ ) of sequences with respect to 14P2 ( $\alpha_{ACP}$  of 14P2 =

1.58 mJ m<sup>-2</sup>, A14P2 = 1.1 × 10<sup>10</sup>) for each sequence vs. ratio of net charge of sequences with respect to 14P2 (-1.4) at pH 7.4. In all cases, the values of  $\sigma_{ACP}$  used here are based on activity product at which growth rates extrapolate to zero. See Supplementary Fig. 15 for plots for  $\sigma_{ACP}$  based on the Ksp from a previous study<sup>18</sup> and Statistic section of Methods for error analysis and n.



**Figure 4**

Proposed mechanism for nanoribbon guided mineralization. (a) Cartoon of pre-nucleation calcium phosphate clusters superimposed on stabilized sheet conformation of p14P2 nanoribbon shows potential for epitaxial match. (b) Proposed role of phosphorylated Amel NR; NR lower the barrier ( $\Delta G$ ) for nucleation of ACP through binding of pre-nucleation complexes on 14P2 domain and following a multi-step crystallization pathway. Phosphorylated 14P2 domain induces ACP to nucleate and spread laterally. ACP eventually transforms to a crystalline phase (OCP or AP), followed by binding with the C-terminus cleaved by Matrix metalloproteinase-20 (MMP20). The template structure is retained, the crystal continues to grow along the c-axis, and NR are finally enzymatically digested by Kallikrein-4 (KLK-4).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [210220v6.1supplementaryNatMat.pdf](#)
- [MovieS114P2ss0.04x387.mov](#)
- [MovieS214P2Ctermss0.04x422.mov](#)
- [MovieS3rh174ss0.04x445.mov](#)

- [MovieS4p14p2ss0.04x464.mov](#)
- [MovieS5p14P2HAPss0.04x437.mov](#)
- [MovieS6p14p2ss0.13x504.mov](#)