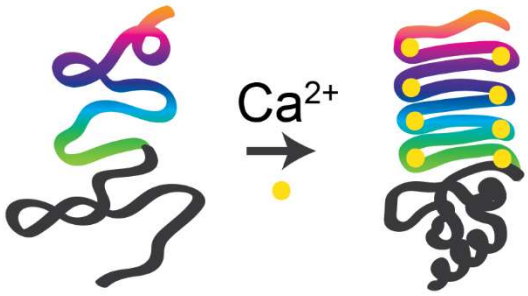
Sequence-defined structural transitions by calcium-responsive proteins

Marina P. Chang,[1] Winnie Huang,[2] Gatha M. Shambharkar,[1] Kenny M. Hernandez,[2] and Danielle J. Mai*[,1,2]

[1]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305
[2]Department of Chemical Engineering, Stanford University, Stanford, CA

*To whom correspondence may be addressed. Email: djmai@stanford.edu



Summary: Recombinant protein engineering accelerated the synthesis of sequence-defined biopolymers that undergo calcium-responsive conformational changes. A mutation panel of repeats-in-toxin (RTX) proteins revealed sequence-dependent disorder, calcium sensitivity, and structural transitions.

Abstract

Biopolymer sequences dictate their functions, and protein-based polymers are a promising platform to establish sequence–function relationships for novel biopolymers. To efficiently explore vast sequence spaces of natural proteins, sequence repetition is a common strategy to tune and amplify specific functions. This strategy is applied to repeats-in-toxin (RTX) proteins with calcium-responsive folding behavior, which stems from tandem repeats of the nonapeptide GGXGXDXUX in which X can be any amino acid and U is a hydrophobic amino acid. To determine the functional range of this nonapeptide, we modified a naturally occurring RTX protein that forms β-roll structures in the presence of calcium. Sequence modifications focused on calcium-binding turns within the repetitive region, including either global substitution of nonconserved residues or complete replacement with tandem repeats of a consensus nonapeptide GGAGXDTLY. Some sequence modifications disrupted the typical transition from intrinsically disordered random coils to folded β rolls, despite conservation of the underlying nonapeptide sequence. Proteins enriched in smaller, hydrophobic amino acids adopted secondary structures in the absence of calcium and underwent structural rearrangement in calcium-rich environments. In contrast, proteins with bulkier, hydrophilic amino acids maintained intrinsic disorder in the absence of calcium. These results indicate a significant role of nonconserved amino acids in calcium-responsive folding, thereby revealing a strategy to leverage sequence in the design of tunable, calcium-responsive biopolymers.

Introduction

Defining the sequence of a polymer is a powerful approach to tune intramolecular conformations, intermolecular interactions, and material properties (1-3). Sequence-defined polymers have enhanced control over self-assembled structures (4-6), molecular recognition (7-9), and stimuli-responsive functions (10-12). As synthetic strategies for sequence-defined polymers continue to improve, tradeoffs emerge between exhaustive or efficient exploration of expansive design spaces (13-16). Such tradeoffs are mitigated by evolutionary processes in biological systems, in which genetic drift and selective pressures can produce diverse traits. Natural macromolecules that have evolved to carry out specific functions are promising platforms to evaluate the level of sequence definition required to design functional polymers.

A function of recent interest is the calcium-responsive folding of bacterial proteins, which critically enable cells to secrete pathogens (17-19), assemble pore-forming toxins (20, 21), and crystallize cell-protective surface layers (22, 23). Calcium responsiveness emerges from conserved, repetitive protein sequences. Specifically, the proteins contain tandem repeats of the "consensus" nonapeptide $(GGXGXDXUX)_n$, where X can be any amino acid, U is an aliphatic amino acid, and n is the number of tandem repeats. The consensus nonapeptide is historically named the Repeats-in-Toxin (RTX) motif, but not all RTX-containing proteins are cytotoxic. In the absence of calcium ions, RTX regions adopt intrinsically disordered conformations (24). In the presence of calcium ions, RTX regions form β-roll structures that consist of parallel β sheets connected by calcium-binding turns (**Figure 1A-B**) (25, 26).

The calcium-responsive folding of the RTX motif inspired recent technological advances beyond the context of bacteria (27, 28). These advances leveraged reversible changes in protein size and surface chemistry upon the introduction of calcium. RTX domains enabled switchable mesh sizes in protein networks (29, 30), calcium-induced crosslinking of protein-based hydrogels (31-34),

regulation of biomolecular recognition (35, 36), column-free purification of recombinant proteins (37, 38), and selective binding of lanthanide ions (39, 40). While RTX-based technologies are promising, a potential limitation is the need for relatively high calcium concentrations (1–100 mM) to initiate folding (41). This need reflects the origin of the RTX motif, which folds in response to calcium concentrations that are relevant for bacteria. RTX proteins must remain disordered in intracellular environments, where calcium concentrations are less than 100 nM. Folding is only initiated upon translocation and secretion into extracellular environments, where calcium concentrations range from 10 μM to >10 mM (42, 43).

The calcium binding affinities of RTX domains are sensitive to sequence, despite the conserved pattern underlying the nonapeptide repeats. This sensitivity is apparent in the well-studied adenylate cyclase toxin (CyaA) of *Bordetella pertussis*, which contains 40 RTX nonapeptide repeats that form five distinct blocks (26). Each CyaA block is denoted with roman numerals I to V corresponding to N- and C- terminal domains, respectively. The fifth block of CyaA—hereafter denoted as "Block V"—binds the most strongly to calcium ions. Calcium-responsive folding proceeds successively from the C-terminus to the N-terminus due to weaker affinities of Blocks IV, III, II, and I (19, 44, 45). Block V consists of nine tandem repeats of the RTX consensus nonapeptide and is flanked by a C-terminal capping domain (**Figure 1A**). The capping domain initiates folding upon secretion through the type I secretion system in Gram-negative bacterial cells. Truncation or removal of the capping domain disrupts calcium-responsive folding, which can be recovered by entropic stabilization of the C-terminus (46, 47). The importance of sequence patterning in Block V was demonstrated by rearranging the order of nonapeptide repeats, which reduced calcium binding affinities (48). This finding suggests that the consensus nonapeptide does not fully describe the requirements for calcium-responsive folding of RTX proteins.

In this work, we modified the sequence of Block V to compare the roles of amino acid size, electrostatic interactions, hydrophobicity, and sequence repetition on the calcium-responsive

folding of RTX proteins. We leveraged recombinant protein engineering to generate twelve sequence variants and systematically evaluate sequence-dependent secondary structures in the absence and presence of calcium. Many sequence variants formed secondary structures in the absence of calcium, in contrast to the intrinsically disordered Block V. Generally, sequence variants exhibited weaker calcium-responsive folding than Block V. Sequence variants that maintained disordered conformations in the absence of calcium underwent weaker folding transitions, revealing the importance of residue size and hydrophobicity in frustrating secondary structure formation. Sequence variants that adopted secondary structures in the absence of calcium underwent calcium-responsive structural rearrangement, revealing unexpected transitions between helical and sheet-like structures. The consistent calcium-bound structures of highly repetitive sequence variants suggest the importance of nonconserved residues in the final folded state of RTX proteins.

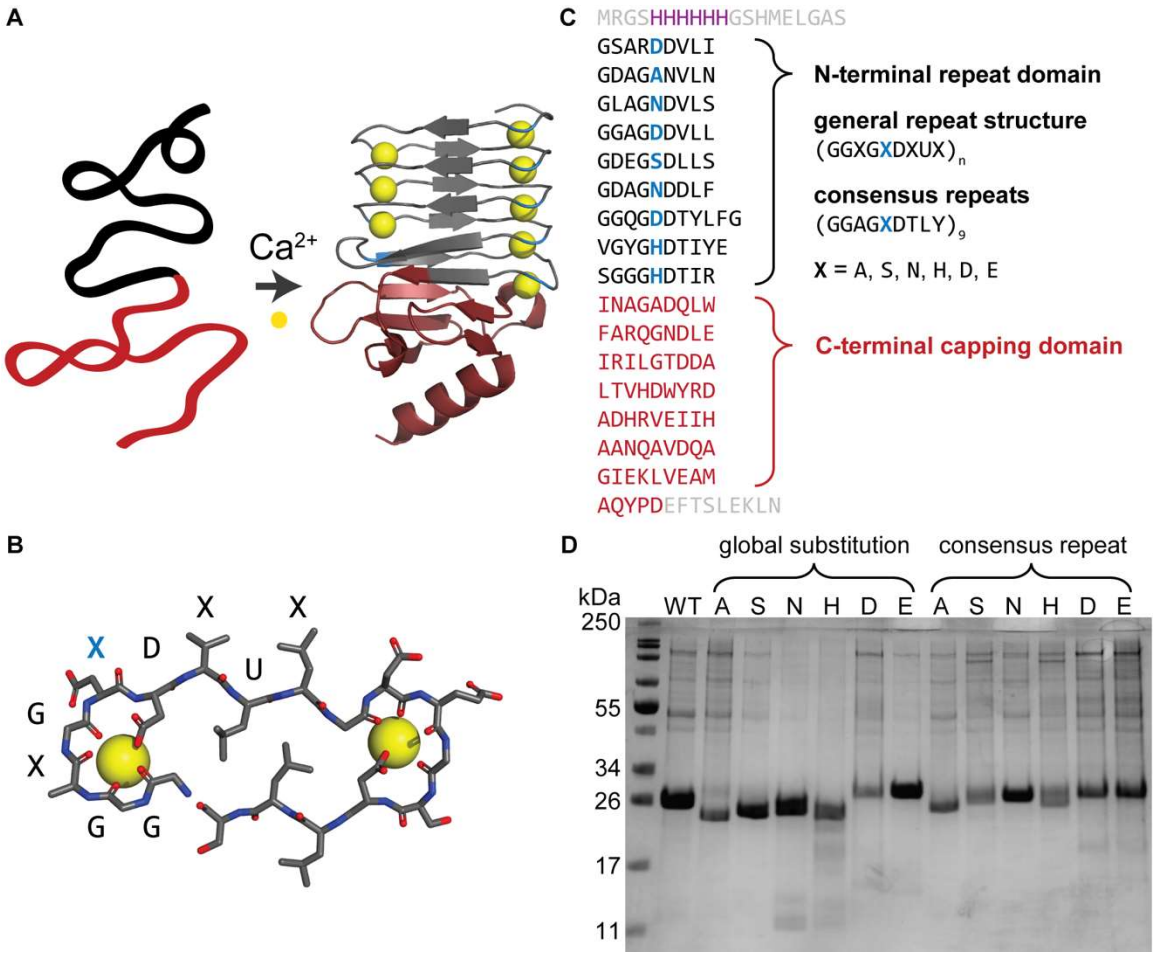


Figure 1. RTX sequence variants were designed to screen the importance of sequence conservation, residue size, hydrophobicity, and electrostatics on calcium-responsive folding. (A) The RTX protein comprises an N-terminal domain (top, black) that is highly repetitive and a C-terminal capping domain (bottom, red) that initiates folding in response to calcium ions (yellow). (B) Top-down view of calcium-binding turns connected by beta sheets. The N-terminal domain is characterized by the repeat sequence GGXGXDXUX, where X indicates a variable amino acid and U is an aliphatic amino acid (PDB: 5CVW (18)), structures generated in Pymol (49).) (C) Primary sequence of CyaA Block V and the C-terminal capping domain prior to mutation. In global substitution variants, all blue residues were replaced with the same amino acid. In consensus repeat variants, the entire N-terminal domain was replaced with 9 tandem repeats of the consensus sequence GGAG**X**DTLY (37). The C-terminal domain was preserved in all sequence variants. Expressed proteins carried additional residues from the directional cloning strategy (gray), as well as a 6×His tag for purification (purple). (D) Recombinant protein expression in *E. coli* was tolerant to all designed mutations, as demonstrated by sodium dodecyl sulfate polyacrylamide gel electrophoresis of purified RTX sequence variants (12% polyacrylamide, 200 V, 45 minutes).

Experimental

Design of RTX sequence variants. To determine the role of sequence in calcium-responsive folding, we produced twelve RTX sequence variants with modifications of the repeat domain Block V (**Figure 1C**). All sequence variants preserved the native C-terminal capping domain to stabilize calcium-bound structures (46, 47). One subset of sequence variants included global substitutions of the nonconserved residue **X** in the fifth position of the nonapeptide GGXG**X**DXUX, which was selected for its proximity to the highly conserved aspartic acid residue in the calcium-binding turn. Global substitution variants replaced nine residues throughout the Block V sequence with a single amino acid. The fifth position of each nonapeptide was globally replaced with either alanine, histidine, serine, asparagine, aspartic acid, or glutamic acid—these options include the five amino acids that occur naturally in these positions throughout Block V, as well as glutamic acid for its chemical similarity to aspartic acid and potential to interact with calcium ions. Another subset of sequence variants replaced Block V with the minimal consensus sequence GGAG**X**DTLY, which was derived from the most common amino acids in a set of RTX-containing proteins (37). Consensus repeat variants included nine tandem repeats of the minimal consensus sequence to match the size of Block V. For each of the six consensus repeat variants, the fifth position **X** included one of the same six amino acids as the global substitution variants: alanine, histidine, serine, asparagine, aspartic acid, or glutamic acid. Complete amino acid sequences and DNA sequences are included in the Supporting Information.

Cloning. Genes encoding Block V and its 12 sequence variants were produced using directional cloning. Genes for each RTX sequence variant were flanked with restriction sites for directional cloning, codon optimized for *Escherichia coli* with scrambling to suppress recombination of repetitive regions (50), and purchased as gene fragments (Twist Bioscience). Genes were subcloned into pQE-9 using BamHI and HindIII restriction sites. All cloning was performed in NEB 5-alpha *E. coli* (New England Biolabs), which were prepared as chemically competent cells using

Mix & Go transformation kits (Zymo Research). Plasmid DNA was purified by miniprep (ZymoPURE) to screen successful cloning through analytical digests at XbaI and SacI restriction sites prior to Sanger sequencing of the inserted region (GENEWIZ). All plasmids are available for use from Addgene.

Protein expression. RTX sequence variants were produced using recombinant protein expression in *E. coli* (51). All expression strains were purchased from New England Biolabs and prepared as chemically competent cells using Mix & Go transformation kits (Zymo Research). Most sequence variants were expressed in T7 Express *lysY*/I^q (NEB), with the exceptions of Block V in T7 Express, alanine global substitution in BL21(DE3), alanine, aspartic acid, and glutamic acid consensus repeats in BL21, and histidine and serine consensus repeats in NEBExpress I^q. Proteins were expressed by inoculating 10 mL of freshly grown overnight culture into 1 L LB media supplemented with 100 µg/mL ampicillin. Cultures were incubated at 37 °C until reaching an optical density at 600 nm between 0.8–1.0. Expression was induced with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG), and expression proceeded for 6 hours at 37 °C. Cells were harvested by centrifugation at 4000 rpm for 10 minutes. Pelleted cells were resuspended in 25 mL of denaturing lysis buffer (100 mM sodium phosphate, 10 mM Tris, 8 M urea, pH 8.0) and stored at –80 °C. To improve yield, lysis buffers for some expressions were supplemented with 1.0 M NaCl (52).

Protein recovery, purification, and validation. Expressed proteins were recovered from cell pellets prior to isolation using immobilized metal affinity chromatography to capture 6×His-tagged proteins of interest, dialysis to remove excess ions, and lyophilization to remove water. To aid defrosting of cell pellets, an additional 25 mL lysis buffer supplemented with 20 mM imidazole was added prior to lysis by sonication. Crude lysates were clarified by centrifugation (8,000 rpm for 1 hour) and filtration (0.45 µm). 6×His-tagged proteins were isolated using immobilized affinity chromatography, in which clarified lysates were incubated with HisPur™ Ni-NTA resin

(ThermoScientific) for 2 hours at ambient temperature. Protein-bound resins were washed with lysis buffer supplemented with 10 mM to 25 mM imidazole prior to elution in lysis buffer supplemented with 250 mM imidazole. Eluted fractions were dialyzed against a chelating buffer (10 mM Tris, 1 mM EGTA, 50 mM NaCl, pH 8.0, 3 exchanges) and ultrapure water (18.2 MΩ·cm, MilliQ, 7 exchanges). Water was removed by lyophilizatio(36)n, and purified proteins were stored at –20 °C. Typical protein expression yields ranged from 14–130 mg per 1 L culture. Protein purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (**Figure 1D**). Protein identity was confirmed with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS, Bruker Microflex LRF) by comparing the measured molar mass to the expected molar mass based on amino acid sequence (Table S1).

Circular dichroism (CD) spectroscopy. Sequence-dependent and calcium-responsive structural changes were measured using CD spectroscopy. Lyophilized proteins were resuspended in 50 mM Tris (pH 7.5) supplemented with up to 100 mM $CaCl_2$ at final protein concentrations between 5 μM and 10 μM. Concentrations were measured after filtration (0.2 μm polyethersulfone membrane) using UV-vis spectroscopy (see SI for details). CD experiments were conducted using a Jasco J-815 Spectropolarimeter. Samples were loaded into a 1 mm pathlength cuvette (Hellma) and held at 20 °C. Scans were performed from 250 nm to 190 nm with 0.2 nm steps and 2 s integration times. Spectra were averaged between 10 scans, and triplicate solutions were measured for each sequence variant. All spectra were corrected by background subtraction of 50 mM Tris (pH 7.5) with the corresponding concentration of $CaCl_2$.

Results and Discussion

Emergence of structure in RTX sequence variants without calcium. In the absence of calcium ions, RTX proteins are typically disordered; however, several RTX sequence variants formed secondary structures that were characterized using CD spectroscopy (**Figure 2A**). In Block V, disorder was indicated by a prominent negative peak at 200 nm, consistent with random coil conformations (24). This peak persisted in global substitution variants with histidine, asparagine, and glutamic acid, which have bulky side chains that promote disorder. In the Block V sequence, these three amino acids each appear multiple times in the nonconserved position of interest. Interestingly, the asparagine variant was nearly indistinguishable from Block V. The reduced intensity of the peak in the histidine variant was attributed to UV absorption by aromatic side chains (53).

In contrast, global substitution variants with alanine, serine, and aspartic acid formed more ordered secondary structures without calcium, suggesting influences of amino acid size and electrostatic interactions. For these variants, the negative peak at 200 nm was replaced by a lower intensity negative peak between 205 nm and 208 nm, and a broad feature from 215 nm to 230 nm appeared. These spectral features suggested the formation of helical structures, which are commonly associated with negative peaks at 208 nm and 222 nm (54). The relative helical content between these variants contradicted the typical helix propensities of alanine (highest), serine (moderate), and aspartic acid (low), indicating the possible role of electrostatic stabilization in secondary structure formation (55-57).

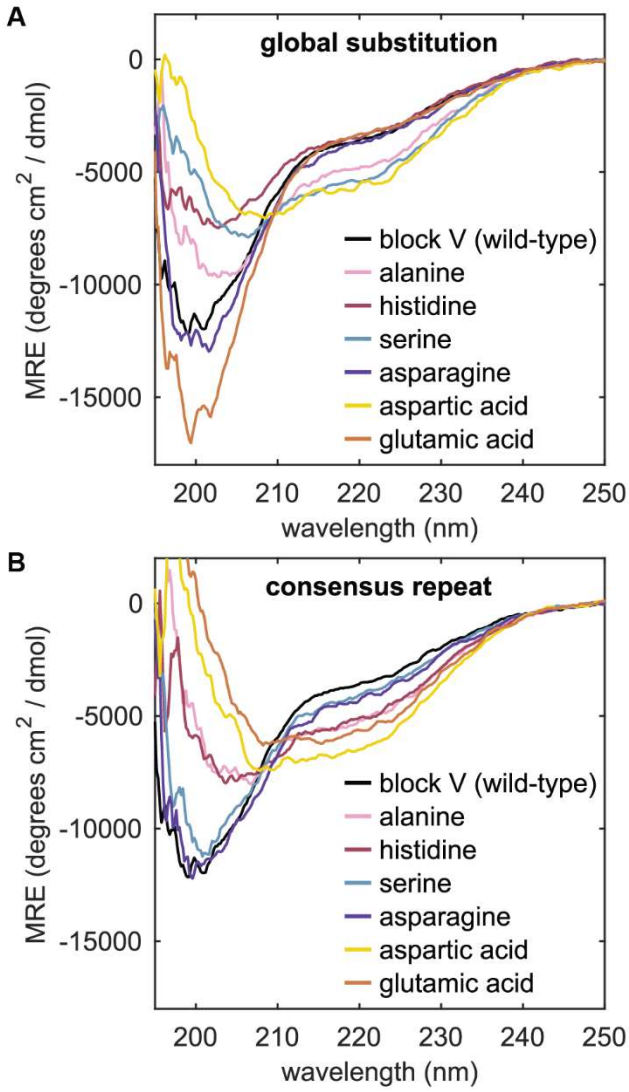


Figure 2. RTX sequence variants adopted diverse secondary structures in the absence of calcium. (A) Secondary structures emerged in CD spectra of global substitution variants with alanine, serine, and aspartic acid. Histidine, asparagine, and glutamic acid variants were disordered and resembled Block V, as indicated by a negative peak in molar residue ellipticity (MRE) at 200 nm. (B) Secondary structures emerged in CD spectra of consensus repeat variants with alanine, histidine, aspartic acid, and glutamic acid. Serine and asparagine variants were disordered and resembled Block V. Replicate spectra for all sequence variants are included in Figures S1–S13.

Among the consensus repeat variants, polar variants mimicked the random coil conformations of Block V without calcium, whereas hydrophobic and charged variants formed more ordered secondary structures (**Figure 2B**). Consensus repeat variants with serine and asparagine formed disordered structures that most resembled Block V, similar to the global substitution asparagine

variant. These disordered structures indicate that polar uncharged residues promote random coil conformations in the absence of calcium. Meanwhile, hydrophobic consensus repeat variants with alanine and histidine adopted similar structures to the global substitution variant with alanine, namely a low intensity negative peak between 205 and 208 nm and a broad feature from 215 to 230 nm. Consensus repeat variants with aspartic acid and glutamic acid produced a broad negative feature from 210 to 230 nm, suggesting both helical and β-sheet characteristics. The diverse secondary structures formed by RTX sequence variants suggest an interplay between steric, hydrophilic, and electrostatic contributions to promote random coil conformations in the absence of calcium.

Global substitution variants alter and weaken calcium-responsive folding. CD spectra of Block V revealed a calcium-dependent structural transition, consistent with prior reports of RTX proteins (**Figure 3A**) (24, 26, 41, 46). Below 0.5 mM $CaCl_2$, Block V adopted a random coil conformation indicated by the negative peak at 200 nm. Above 0.5 mM $CaCl_2$, Block V formed β-sheet structures indicated by the appearance of a negative peak at 218 nm and the disappearance of the negative peak at 200 nm. Deconvolution of CD spectra taken at 0 and 100 mM $CaCl_2$ revealed an increase in sheet content from 19.1% to 27.9% upon the addition of calcium ions (**Figure 3B**), which is denoted as a 46% relative increase in sheet content. Block V also produced a 59% relative increase in turn content, consistent with the formation of the β-roll structure characteristic of RTX proteins. A modest 17% relative increase in helical content was attributed to folding of the capping domain. Spectral deconvolution was performed from 200 nm to 250 nm with CDPro Software using the reference set SPD48, which is the largest available reference set that includes denatured proteins (58). The results from the CDSSTR, CONTIN/LL, and SELCON3 methods were normalized and averaged to facilitate quantitative comparisons.

Many of the global substitution variants underwent calcium-responsive structural changes that were not characteristic of β-roll formation (**Figure 4A**, top row). These unexpected structural

rearrangements required higher calcium concentrations than those of Block V. For alanine, serine, and aspartic acid variants, the addition of 10 mM $CaCl_2$ corresponded with disappearance of the negative peak between 205 nm and 208 nm. The three variants produced distinct changes in the broad feature from 215 nm to 230 nm, such that the feature was enhanced for the alanine variant, relatively constant for the serine variant, and reduced for the aspartic acid variant. For all three variants, spectral deconvolution indicated ≤10% relative increases in sheet content from 0 mM to 100 mM $CaCl_2$ (**Figure 4B**).

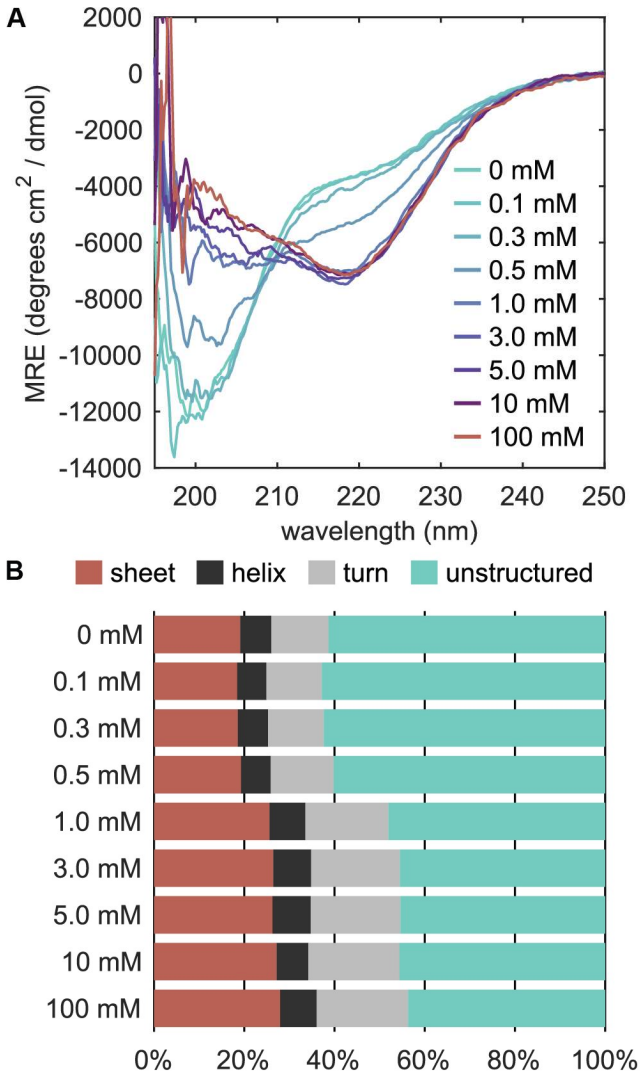


Figure 3. Block V formed β-roll structures in the presence of calcium ions. (A) CD spectroscopy of Block V revealed a transition from disordered random coils to β-roll structures between 0.5 mM

and 1.0 mM $CaCl_2$. Replicate spectra are included in Figure S1. (B) Spectral deconvolution quantified a structural transition at 1.0 mM $CaCl_2$. This transition produced increases in sheet, helix, and turn content and a corresponding decrease in unstructured content.

Disordered global substitution variants with histidine, asparagine, and glutamic acid formed β-roll structures upon addition of sufficient calcium chloride (**Figure 4A**, bottom row). For the histidine variant, addition of 5 mM $CaCl_2$ corresponded with disappearance of the negative peak at 200 nm and the appearance of a low-intensity negative peak at 225 nm. Spectral deconvolution indicated 35% and 48% relative increases in sheet and turn content between 0 and 100 mM $CaCl_2$ (**Figure 4B**), suggesting that UV absorbance by histidine obscured the typical signatures of β-roll formation. For the asparagine variant, a sharp transition near 1.0 mM $CaCl_2$ resembled the Block V transition at 0.5 mM $CaCl_2$. For the glutamic acid variant, folding occurred gradually from 5 mM to 100 mM $CaCl_2$.

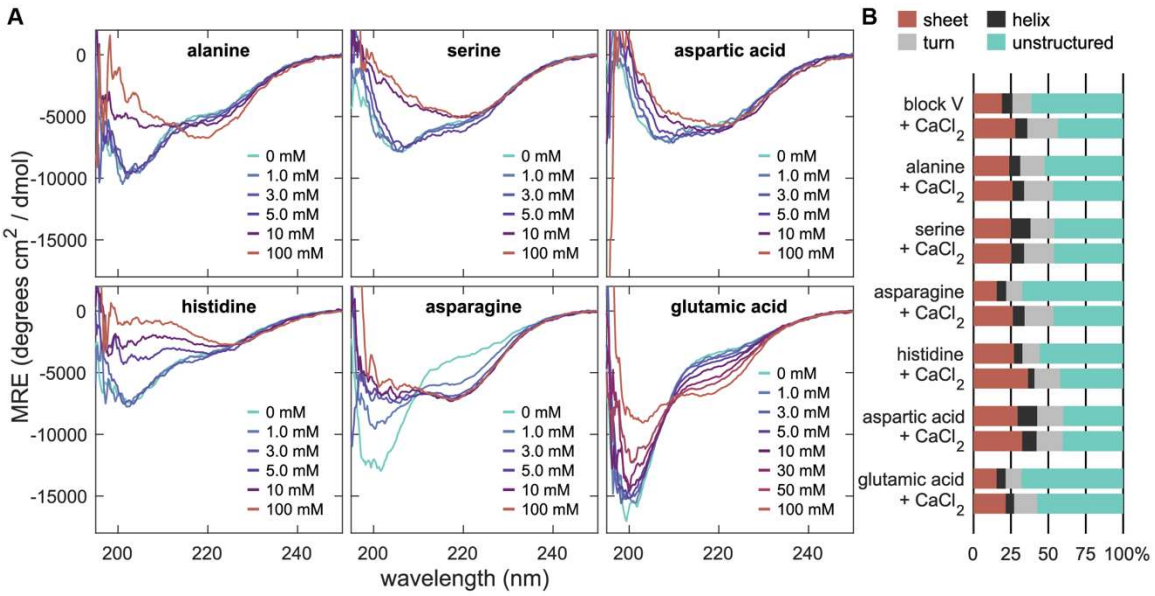


Figure 4. Calcium-responsive folding of global substitution variants was weaker than Block V. (A) Asparagine and glutamic acid variants formed the most similar calcium-responsive structures to Block V, whereas alanine, serine, histidine, and aspartic acid variants underwent qualitatively different structural changes. Replicate spectra for all global substitution variants are included in Figures S2–S7. (B) CD spectral deconvolution at 0 mM and 100 mM $CaCl_2$ revealed that the greatest secondary structure changes in response to calcium occurred in variants that were most disordered without calcium.

Table 1. Global substitutions of Block V reduce binding to calcium

Variant	K_D (mM)	n
Block V	0.67 ± 0.08	4.1 ± 1.3
Asparagine	1.0 ± 0.8	1.8 ± 0.7
Glutamic acid	11 ± 3	1.0 ± 0.3

The global substitution variants with asparagine and glutamic acid demonstrated weaker calcium affinity and reduced cooperativity compared to Block V (**Table 1**). To compare the calcium responsiveness of RTX sequence variants to Block V, the Hill–Langmuir equation was used to fit the fraction of protein bound by calcium ions θ with respect to total calcium concentration $[\mathrm{Ca}^{2+}]$ (59, 60):

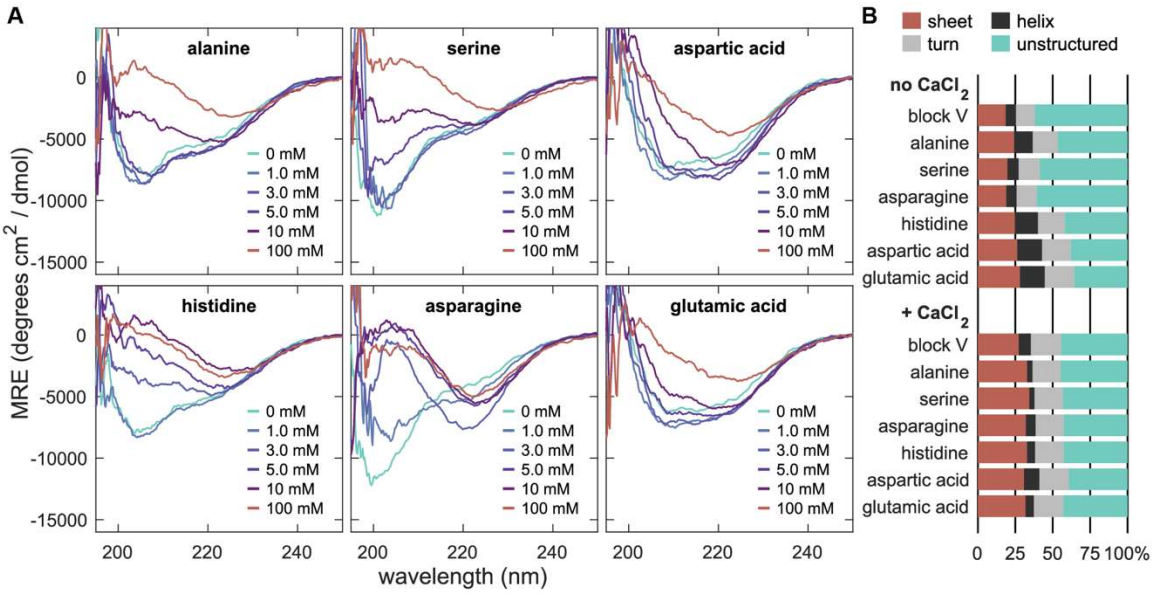
$$\theta = \frac{[\mathrm{Ca}^{2+}]^n}{K_D^n + [\mathrm{Ca}^{2+}]^n}$$

where the Hill coefficient n describes the cooperativity of ligand binding, and the half-saturation dissociation constant K_D indicates the calcium concentration at which half of protein binding sites are occupied. θ was calculated by normalizing the molar residue ellipticity at 218 nm, with the maximum absolute intensity corresponding to complete binding and β-roll formation. The asparagine variant resembled Block V, with similar K_D and positively cooperative binding ($n > 1$). However, cooperative binding was weaker for the asparagine variant than for Block V. The glutamic acid variant exhibited an order-of-magnitude weaker response to calcium, which may result from its noncooperative binding. This analysis was limited to the asparagine and glutamic acid variants, which maintained disordered structures in the absence of calcium and produced the characteristic β-roll signature at 218 nm. The calcium-free structures of the remaining global substitution variants prevented reliable quantification of θ.

The weaker calcium-responsive folding of global substitution variants than Block V emphasizes the importance of sequence evolution in natural RTX proteins. The consensus RTX sequence

GGXGXDXUX highlights some necessary features for RTX proteins to function, such as glycine for flexibility in the calcium-binding turn, aspartic acid to stabilize electrostatic interactions of divalent cations, and aliphatic residues to form the characteristic β-roll structure (25). However, these features alone were not sufficient to facilitate calcium-responsive folding of RTX sequence variants. Sequence variants that formed secondary structures in the absence of calcium suggest that the fifth residue of GGXG**X**DXUX plays a role in frustrating protein folding. Frustrated proteins that adopt random coil conformations may sample a broader folding energy landscape that promotes ion-driven folding, whereas proteins with less frustration may fold prematurely into conformations with less favorable ionic interactions (61, 62). This contrast is best highlighted by comparing the variants with aspartic acid and glutamic acid, which have identical net charges but different side chain lengths. This subtle difference in residue structure led to drastically different calcium-responsive structural changes between these variants. The smaller aspartic acid residue promoted electrostatic stabilization of secondary structures, whereas the bulkier glutamic acid residue promoted frustration and random coil conformations. In Block V, strong cooperative binding likely results from a mix of both stabilizing and frustrating residues throughout the repeat domain.

Consensus repeat variants undergo structural rearrangement to form consistent calcium-bound structures. Despite a range of secondary structures in the absence of calcium, all consensus repeat variants adopted similar secondary structures in the presence of 100 mM $CaCl_2$ (**Figure 5A**). CD spectra showed monotonic decreases in ellipticity from 200 nm to 220 nm to produce strong negative peaks near 225 nm (red curves). Spectral deconvolution revealed similar structural components, with calcium-bound structures demonstrating less variation than calcium-free structures (**Figure 5B**). Calcium-bound consensus repeat variants also formed 10–23% relatively higher sheet content compared to calcium-bound Block V.



317

318 **Figure 5.** Consensus repeat variants formed consistent calcium-bound structures. (A) All
319 consensus repeat variants produced similar circular dichroism spectra at 100 mM $CaCl_2$ (red
320 curves), which were characterized by monotonic decreases from 200 nm to 220 nm and a broad
321 negative peak near 225 nm. Replicate spectra for all consensus repeat variants are included in
322 Figures S8–S13. (B) CD spectral deconvolution at 0 mM and 100 mM $CaCl_2$ revealed structural
323 variation among consensus repeat variants in the absence of calcium, in contrast to quantitatively
324 similar structures in the presence of 100 mM $CaCl_2$.

325 Some consensus repeat variants underwent conformational changes from random coils to β-roll
326 structures, whereas others underwent calcium-responsive structural rearrangements. In the
327 absence of calcium, consensus repeat variants with serine and asparagine maintained the most
328 disorder. The serine and asparagine variants underwent characteristic RTX folding transitions,
329 respectively showing 26% and 30% relative decreases in unstructured content between 0 mM
330 and 100 mM $CaCl_2$. Meanwhile, alanine, histidine, aspartic acid, and glutamic acid variants
331 revealed unexpected structural transitions in response to calcium. For these variants, increases
332 in sheet content were associated with decreases in helical content, resulting in the unstructured
333 content remaining similar in the absence and presence of calcium for the alanine, histidine, and
334 aspartic acid variants. The glutamic acid variant produced a 20% relative increase in unstructured
335 content between 0 mM and 100 mM $CaCl_2$. Interestingly, these calcium-responsive changes in

secondary structure revealed transitions between helical and sheet-like structures that are unlike the conformational changes from random coils to β-rolls by Block V.

Like global substitution variants, consensus repeat variants demonstrated weaker sensitivity to calcium compared to Block V. The consensus repeat variant with asparagine retained the greatest sensitivity, with conformational changes occurring between 1.0 mM and 3.0 mM $CaCl_2$. For serine and histidine variants, structural transitions occurred gradually between 3.0 mM and 100 mM $CaCl_2$. The alanine, aspartic acid, and glutamic acid variants exhibited the weakest calcium sensitivities, with structural transitions occurring between 10 mM and 100 mM $CaCl_2$. The reduced calcium sensitivity of all sequence variants in this work suggests that nonconserved residues and sequence patterns are necessary to maintain the calcium sensitivity of Block V.

General Conclusions. There remains much to learn from nature's design rules for calcium-responsive protein folding. To probe sequence effects, we modified the repetitive region of Block V—a naturally occurring RTX protein domain that binds to calcium by folding into a parallel β-roll. Global substitution variants altered the size, charge, and hydrophobicity of nonconserved residues in the calcium-binding turns of Block V, and consensus repeat variants replaced the repetitive region of Block V with tandem repeats of the nonapeptide GGAGXDTLY. All sequence mutations were tolerated during recombinant protein expression, which accelerates the rapid and accurate production of sequence-defined biopolymers.

Despite changes to nonconserved residues, RTX variants adopted diverse, sequence-dependent secondary structures ranging from random coil conformations resembling Block V to more helical structures. In the global substitution variants, random coil conformations were achieved by the largest residues: histidine, asparagine, and glutamic acid. Unanticipated helical structures were observed for the global substitution variants with the smallest residues, alanine and serine. Residue size effects were further emphasized by unexpected helical structures formed by variants

with aspartic acid, which contrasted the random coil conformations of variants with glutamic acid. For the consensus repeat variants, the hydrophilic residues serine and asparagine most resembled Block V in the absence of calcium. In the nonconserved position of interest, bulkier and hydrophilic residues tended to frustrate protein folding, enabling the protein to maintain a disordered structure in the absence of calcium.

RTX sequence variants that preserved intrinsic disorder in the absence of calcium underwent calcium-responsive folding transitions associated with β-roll formation. β-roll structures emerged for global substitution variants with histidine, asparagine, and glutamic acid, although each with a weaker calcium affinity and cooperativity than Block V. Consensus repeat variants with polar residues—serine and asparagine—also underwent calcium-responsive folding. In contrast, sequence variants that adopted secondary structures in the absence of calcium revealed calcium-responsive structural rearrangements, in which increases in sheet content were offset by decreases in helical content. These transitions appear unlike the characteristic folding of random coils into β rolls by natural RTX proteins. Moreover, consensus repeat variants adopted different final structures than Block V, specifically with higher sheet content in the presence of 100 mM $CaCl_2$.

Overall, our results highlight the versatility of recombinant protein engineering to map sequence–function relationships of biopolymers. We establish the importance of size and hydrophobicity of nonconserved residues in the RTX nonapeptide GGXGXDXUX. Asparagine strikes a particular balance between size and hydrophilic character, demonstrating the most calcium sensitivity within the sets of global substitution and consensus repeat variants. We anticipate these insights will advance the use of RTX proteins as tunable, ion-responsive components of protein-based biomaterials and biotechnologies.

Author Contributions: M.P.C. and D.J.M. conceptualized the study and designed experiments. M.P.C., W.H., G.M.S., and K.M.H. conducted molecular cloning, expressed and purified recombinant proteins, and validated expressed proteins. M.P.C. conducted circular dichroism measurements. M.P.C. and D.J.M. analyzed data. M.P.C. and D.J.M. wrote the initial draft of the manuscript. All authors contributed to the revision and editing process. D.J.M. supervised the research.

Acknowledgments

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-22-1-0241. This work was supported in part by the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, under Contract No. DE-AC02-76SF00515 with the U.S. Department of Energy, Office of Basic Energy Sciences. We recognize support from the Department of Chemical Engineering, Bio-X Summer Undergraduate Research Program, and Office of the Vice Provost for Undergraduate Education at Stanford University. MALDI-TOF-MS measurements were supported by the Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University Mass Spectrometry (RRID:SCR_017801) utilizing the Bruker Microflex MALDI TOF mass spectrometer (RRID:SCR_018696). We thank Prof. Bradley Olsen for the generous gift of a pQE-9-mCherry plasmid, Prof. Possu Huang and Carla Perez for CD spectroscopy access, and Alana Gudinas and all members of the Mai Lab for helpful conversations.

References

1. J.-F. Lutz, The future of sequence-defined polymers. *European Polymer Journal* **199**, 112465 (2023).
2. A. J. DeStefano, R. A. Segalman, E. C. Davidson, Where Biology and Traditional Polymers Meet: The Potential of Associating Sequence-Defined Polymers for Materials Science. *JACS Au* **1**, 1556-1571 (2021).
3. S. L. Perry, C. E. Sing, 100th Anniversary of Macromolecular Science Viewpoint: Opportunities in the Physics of Sequence-Defined Polymers. *ACS Macro Letters* **9**, 216-225 (2020).

4. H. Yu *et al.*, Self-Assembly of Repetitive Segment and Random Segment Polymer Architectures. *ACS Macro Letters* **11**, 1366-1372 (2022).
5. J. Park, A. Staiger, S. Mecking, K. I. Winey, Ordered Nanostructures in Thin Films of Precise Ion-Containing Multiblock Copolymers. *ACS Central Science* **8**, 388-393 (2022).
6. J. Babi *et al.*, Self-assembled free-floating nanomaterials from sequence-defined polymers. *Journal of Polymer Science* **59**, 2378-2404 (2021).
7. M. Szatko, W. Forysiak, S. Kozub, T. Andruniów, R. Szweda, Revealing the Effect of Stereocontrol on Intermolecular Interactions between Abiotic, Sequence-Defined Polyurethanes and a Ligand. *ACS Biomaterials Science & Engineering* **10**, 3727-3738 (2024).
8. B. M. Seifried *et al.*, Glycoprotein Mimics with Tunable Functionalization through Global Amino Acid Substitution and Copper Click Chemistry. *Bioconjugate Chemistry* **31**, 554-566 (2020).
9. S. Celasun *et al.*, Digging into the Sequential Space of Thiolactone Precision Polymers: A Combinatorial Strategy to Identify Functional Domains. *Angewandte Chemie International Edition* **58**, 1960-1964 (2019).
10. B. M. Wirtz, A. G. Yun, C. Wick, X. J. Gao, D. J. Mai, Protease-Driven Phase Separation of Elastin-Like Polypeptides. *Biomacromolecules* **25**, 4898-4904 (2024).
11. X. Yuan, H. W. Hatch, J. C. Conrad, A. B. Marciel, J. C. Palmer, pH response of sequence-controlled polyampholyte brushes. *Soft Matter* **19**, 4333-4344 (2023).
12. C. Pan *et al.*, Rewritable two-dimensional DNA-based data storage with machine learning reconstruction. *Nature Communications* **13**, 2984 (2022).
13. M. A. Webb, N. E. Jackson, P. S. Gil, J. J. de Pablo, Targeted sequence design within the coarse-grained polymer genome. *Science Advances* **6**, eabc6216.
14. R. A. Patel, C. H. Borca, M. A. Webb, Featurization strategies for polymer sequence or composition design by machine learning. *Molecular Systems Design & Engineering* **7**, 661-676 (2022).
15. P. S. Ramesh, T. K. Patra, Polymer sequence design via molecular simulation-based active learning. *Soft Matter* **19**, 282-294 (2023).
16. E. C. Day, S. S. Chittari, M. P. Bogen, A. S. Knight, Navigating the Expansive Landscapes of Soft Materials: A User Guide for High-Throughput Workflows. *ACS Polymers Au* **3**, 406-427 (2023).
17. I. Linhartová *et al.*, RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiology Reviews* **34**, 1076-1112 (2010).
18. L. Bumba *et al.*, Calcium-Driven Folding of RTX Domain β-Rolls Ratchets Translocation of RTX Proteins through Type I Secretion Ducts. *Molecular Cell* **62**, 47-62 (2016).
19. L. Motlova, N. Klimova, R. Fiser, P. Sebo, L. Bumba, Continuous Assembly of β-Roll Structures Is Implicated in the Type I-Dependent Secretion of Large Repeat-in-Toxins (RTX) Proteins. *Journal of Molecular Biology* **432**, 5696-5710 (2020).
20. A. J. Wallace *et al.*, *E. coli* Hemolysin E (HlyE, ClyA, SheA): X-Ray Crystal Structure of the Toxin and Observation of Membrane Pores by Electron Microscopy. *Cell* **100**, 265-276 (2000).
21. M. D. Peraro, F. G. van der Goot, Pore-forming toxins: ancient, but never really out of fashion. *Nature Reviews Microbiology* **14**, 77-92 (2016).
22. J. Herrmann *et al.*, Environmental Calcium Controls Alternate Physical States of the Caulobacter Surface Layer. *Biophysical Journal* **112**, 1841-1851 (2017).
23. J. Herrmann *et al.*, A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly. *Proceedings of the National Academy of Sciences* **117**, 388-394 (2020).

24. A. Chenal, J. I. Guijarro, B. Raynal, M. Delepierre, D. Ladant, RTX Calcium Binding Motifs Are Intrinsically Disordered in the Absence of Calcium: Implication for Protein Secretion. *Journal of Biological Chemistry* **284**, 1781-1789 (2009).
25. U. Baumann, S. Wu, K. M. Flaherty, D. B. McKay, Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa: a two-domain protein with a calcium binding parallel beta roll motif. *The EMBO Journal* **12**, 3357-3364 (1993).
26. C. Bauche *et al.*, Structural and Functional Characterization of an Essential RTX Subdomain of Bordetella pertussis Adenylate Cyclase Toxin. *Journal of Biological Chemistry* **281**, 16914-16926 (2006).
27. B. Bulutoglu, S. Banta, Block V RTX Domain of Adenylate Cyclase from Bordetella pertussis: A Conformationally Dynamic Scaffold for Protein Engineering Applications. *Toxins* **9**, 289 (2017).
28. M. P. Chang, W. Huang, D. J. Mai, Monomer-scale design of functional protein polymers using consensus repeat sequences. *Journal of Polymer Science* **59**, 2644-2664 (2021).
29. P. Ringler, G. E. Schulz, Self-Assembly of Proteins into Designed Networks. *Science* **302**, 106-109 (2003).
30. L. Liu, H. Wang, Y. Han, S. Lv, J. Chen, Using single molecule force spectroscopy to facilitate a rational design of Ca2+-responsive β-roll peptide-based hydrogels. *Journal of Materials Chemistry B* **6**, 5303-5312 (2018).
31. K. Dooley, Y. H. Kim, H. D. Lu, R. Tu, S. Banta, Engineering of an Environmentally Responsive Beta Roll Peptide for Use As a Calcium-Dependent Cross-Linking Domain for Peptide Hydrogel Formation. *Biomacromolecules* **13**, 1758-1764 (2012).
32. X.-R. Zhou, R. Ge, S.-Z. Luo, Self-assembly of pH and calcium dual-responsive peptide-amphiphilic hydrogel. *Journal of Peptide Science* **19**, 737-744 (2013).
33. K. Dooley, B. Bulutoglu, S. Banta, Doubling the Cross-Linking Interface of a Rationally Designed Beta Roll Peptide for Calcium-Dependent Proteinaceous Hydrogel Formation. *Biomacromolecules* **15**, 3617-3624 (2014).
34. B. Bulutoglu, S. J. Yang, S. Banta, Conditional Network Assembly and Targeted Protein Retention via Environmentally Responsive, Engineered β-Roll Peptides. *Biomacromolecules* **18**, 2139-2145 (2017).
35. B. Bulutoglu, K. Dooley, G. Szilvay, M. Blenner, S. Banta, Catch and Release: Engineered Allosterically Regulated β-Roll Peptides Enable On/Off Biomolecular Recognition. *ACS Synthetic Biology* **6**, 1732-1741 (2017).
36. W. Abdallah, K. Solanki, S. Banta, Insertion of a Calcium-Responsive β-Roll Domain into a Thermostable Alcohol Dehydrogenase Enables Tunable Control over Cofactor Selectivity. *ACS Catalysis* **8**, 1602-1613 (2018).
37. O. Shur, K. Dooley, M. Blenner, M. Baltimore, S. Banta, A designed, phase changing RTX-based peptide for efficient bioseparations. *BioTechniques* **54**, 197-206 (2013).
38. J. Hendrix *et al.*, Engineered Calcium-Precipitable Restriction Enzyme. *ACS Synthetic Biology* **3**, 969-971 (2014).
39. H. Jung, Z. Su, Y. Inaba, A. C. West, S. Banta, Genetic Modification of Acidithiobacillus ferrooxidans for Rare-Earth Element Recovery under Acidic Conditions. *Environmental Science & Technology* **57**, 19902-19911 (2023).
40. F. Khoury, Z. Su, S. Banta, Rare Earth Element Binding and Recovery by a Beta Roll-Forming RTX Domain. *Inorganic Chemistry* **63**, 13223-13230 (2024).
41. T. Rose, P. Sebo, J. Bellalou, D. Ladant, Interaction of Calcium with *Bordetella pertussis* Adenylate Cyclase Toxin: Characterization of Multiple Calcium-Binding Sites and Calcium-Induced Conformational Changes. *Journal of Biological Chemistry* **270**, 26370-26376 (1995).
42. P. Gangola, B. P. Rosen, Maintenance of intracellular calcium in Escherichia coli. *Journal of Biological Chemistry* **262**, 12570-12574 (1987).

43. M. M. King, B. B. Kayastha, M. J. Franklin, M. A. Patrauchan, "Calcium Regulation of Bacterial Virulence" in Calcium Signaling, M. S. Islam, Ed. (Springer International Publishing, Cham, 2020), 10.1007/978-3-030-12457-1_33, pp. 827-855.
44. H. Wang, X. Gao, H. Li, Single Molecule Force Spectroscopy Reveals the Mechanical Design Governing the Efficient Translocation of the Bacterial Toxin Protein RTX. *Journal of the American Chemical Society* **141**, 20498-20506 (2019).
45. H. Wang, G. Chen, H. Li, Templated folding of the RTX domain of the bacterial toxin adenylate cyclase revealed by single molecule force spectroscopy. *Nature Communications* **13**, 2784 (2022).
46. A.-C. S. Pérez *et al.*, Characterization of the Regions Involved in the Calcium-Induced Folding of the Intrinsically Disordered RTX Motifs from the Bordetella pertussis Adenylate Cyclase Toxin. *Journal of Molecular Biology* **397**, 534-549 (2010).
47. M. A. Blenner, O. Shur, G. R. Szilvay, D. M. Cropek, S. Banta, Calcium-Induced Folding of a Beta Roll Motif Requires C-Terminal Entropic Stabilization. *Journal of Molecular Biology* **400**, 244-256 (2010).
48. O. Shur, S. Banta, Rearranging and concatenating a native RTX domain to understand sequence modularity. *Protein Engineering, Design and Selection* **26**, 171-180 (2012).
49. Schrödinger, LLC., The PyMOL Molecular Graphics System, Version 3.0.
50. N. C. Tang, A. Chilkoti, Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nature Materials* **15**, 419-424 (2016).
51. M. A. Morris *et al.*, Democratizing the rapid screening of protein expression for materials development. *Molecular Systems Design & Engineering* **8**, 227-239 (2023).
52. V. Yeong, E. G. Werth, L. M. Brown, A. C. Obermeyer, Formation of Biomolecular Condensates in Bacteria by Tuning Protein Electrostatics. *ACS Central Science* **6**, 2301-2310 (2020).
53. E. Peggion, A. Cosani, M. Terbojevich, E. Scoffone, Solution Properties of Synthetic Polypeptides. Circular Dichroism Studies on Poly-L-histidine and on Random Copolymers of L-Histidine and L-Lysine in Aqueous Solution. *Macromolecules* **4**, 725-731 (1971).
54. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876-2890 (2006).
55. K. T. O'Neil, W. F. DeGrado, A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science* **250**, 646-651 (1990).
56. M. Blaber, X.-j. Zhang, B. W. Matthews, Structural Basis of Amino Acid α Helix Propensity. *Science* **260**, 1637-1640 (1993).
57. C. Nick Pace, J. Martin Scholtz, A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal* **75**, 422-427 (1998).
58. N. Sreerama, R. W. Woody, Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Analytical Biochemistry* **287**, 252-260 (2000).
59. R. Gesztelyi *et al.*, The Hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences* **66**, 427-438 (2012).
60. D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry* (W. H. Freeman, 2017).
61. A. Kluber, T. A. Burt, C. Clementi, Size and topology modulate the effects of frustration in protein folding. *Proceedings of the National Academy of Sciences* **115**, 9234-9239 (2018).
62. S. Gianni *et al.*, Fuzziness and Frustration in the Energy Landscape of Protein Folding, Function, and Assembly. *Accounts of Chemical Research* **54**, 1251-1259 (2021).