BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

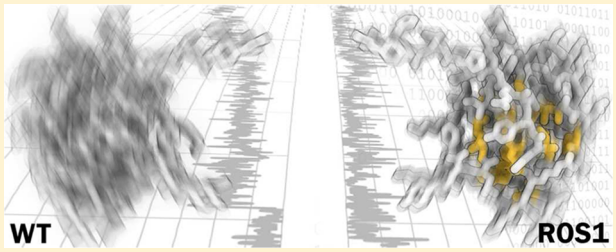
Novel Protease Inhibitors via Computational Redesign of Subtilisin BPN′ Propeptide

Ashley B. Daugherty, Pravin Muthu, and Stefan Lutz*

Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, Georgia 30322, United States

Supporting Information

ABSTRACT: The propeptide domain of subtilisin BPN′ functions as a molecular chaperone for its cognate protease yet quickly assumes a predominantly unfolded structure following cleavage by the mature protease. In contrast, structural stabilization of the propeptide domain has been proposed to competitively inhibit protease self-cleavage, suggesting the possibility for the generation of novel proteinaceous subtilisin inhibitors. Using a Rosetta fixed backbone design, we have redesigned the subtilisin BPN′ propeptide structure to generate synthetic peptide sequences with increased and tunable structural stability. Molecular dynamics simulations provide supporting evidence that the artificial sequences retain structure without its protease cognate unlike the inherently disordered wild-type propeptide. Experimental evaluation of two designer domains by spectroscopic methods verified their structural integrity. Furthermore, the novel propeptide domains were shown to possess significantly enhanced thermostability. Nevertheless, their modest functional performance as protease inhibitors raises doubt that propeptide stability alone is sufficient for effective inhibitor design.



Subtilisins belong to the class of nonspecific serine proteases in the subtilase family. Originally isolated from various *Bacillus* species in soil, these enzymes are highly efficient and promiscuous hydrolases with significance in a variety of commercial applications and organic synthetic chemistry.[1−4] To prevent autoproteolysis of the enzyme, the addition of reversible protease inhibitors such as borates and boronic acid derivatives has been a standard procedure.[5−7] Nevertheless, these small-molecule inhibitors present a significant environmental problem. Attractive and sustainable alternatives to the existing inhibitors are proteinaceous subtilisin inhibitors derived from the protease's own N-terminal propeptide sequence.[4,8] Normally, these propeptide sequences play a critical role as a chaperone in the maturation process of native proteases, assisting the latter in folding efficiently into its correct tertiary structure.[9−11] Upon the successful completion of folding, the propeptide C-terminal region is positioned in the substrate-binding cleft of the mature enzyme and is consequently hydrolyzed. Self-cleavage triggers unfolding and release of the pro-segment, followed by further degradation by its cognate protease.[12−14]

Beyond the function of the propeptide as an intramolecular chaperone in protease maturation, previous studies have indicated that the bound pro-sequence can also temporarily inhibit proteolytic activity.[15,16] A key component to sustaining inhibitory function is retention of the propeptide tertiary structure. It has been reasoned that these naturally effective but short-lived inhibitors can work for extended periods of time if the structural integrity of the propeptide is preserved through enhanced protein stability, which reduces the level of pro-sequence unfolding and susceptibility to proteolysis.[12,17,18] Such an argument is independently supported by structure analysis of natural protease inhibitors such as proteinase A inhibitor 1 (POIA1) from *Pleurotus ostreatus* [Protein Data Bank (PDB) entries 1V5I and 1ITP] and yeast proteinase B inhibitor 2 that have been found to exhibit high degrees of structural similarity with the folded propeptide.[15,19−21]

The hypothesis has also been explored experimentally by a series of protein engineering studies involving subtilisin BPN′ from *Bacillus amyloliquefaciens* and its pro-sequence. Rational protein design and directed evolution strategies led to the introduction of a disulfide bond into the propeptide sequence by Bryan and co-workers and translated into a dramatic stabilization of the propeptide.[22−25] Separately, Kojima et al. reported a number of propeptide BPN′ mutants with improved packing of their hydrophobic core that exhibited increased stability and improved inhibitory function.[17,18,26] However, the former design was not tested for inhibitory function, while the data for the latter variants proved to be difficult to reproduce.

In recent years, advances in computational algorithms for in silico protein design have offered a promising new approach for improving protein stability.[27,28] In silico methods have an advantage over directed evolution strategies with regard to speed and ability for more comprehensively sampling the vastness of sequence space. An extended study of nine globular proteins by Baker and co-workers demonstrated that their Rosetta design algorithm is an effective tool for identifying

Received: June 20, 2012
Revised: September 20, 2012
Published: September 25, 2012

dx.doi.org/10.1021/bi300832v | *Biochemistry* 2012, 51, 8247−8255

amino acid substitutions in proteins and peptides that will enhance their overall thermostability without significant structural perturbations.[29,30] In addition, the arbitrary energy scores assigned to individual designs could offer a convenient, quantitative tool for tuning protein stability. In the case of the redesign of the propeptide into subtilisin inhibitors, such tunability of stability could ideally translate directly into tailoring inhibitory function.

In this paper, we describe the creation of artificial proteinaceous protease inhibitors in silico using the Rosetta design algorithm, as well as their structural evaluation and functional analysis against selected proteases. These latter studies access changes in secondary structure content, thermostability, and inhibitory properties. Furthermore, the design features in the artificial propeptides are probed by site-directed mutagenesis and molecular dynamics simulations to further test and validate the design strategy. Through this study, we are able to provide an alternative tool for the redesign of synthetic propeptides that have increased structural integrity and can be evaluated for their potency as protease inhibitors.

MATERIALS AND METHODS

Materials. Whole-gene synthesis was performed by DNA2.0 (Menlo Park, CA), while individual oligodeoxynucleotides were ordered from Integrated DNA Technologies (Coralville, IA). *Pfu* Turbo DNA polymerase (Stratagene, La Jolla, CA) was used for DNA amplification. Restriction enzymes were purchased from New England Biolabs (Ipswich, MA), while proteases and reagents were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated.

Computational Simulations. The Rosetta Molecular Modeling program (version 3.1) was used for the redesign of the propeptide domain.[31] Modeling was based on the crystallographic coordinates of BPN′ propeptide, complexed with its natural binding partner subtilisin BPN′ (PBD entry 1SPB).[9] We performed three rounds of fixed backbone design, idealizing amino acid sequences to the crystallized backbone torsions.[32] To account for backbone rigidity, amino acid side chain conformations were modeled using an expanded Dunbrack rotamer library for increased flexibility during side chain packing.[33] The first round consisted of 100 independent trajectories, followed by sequence profile analysis using ConSurf.[34] Amino acid substitutions at each position that were observed more than once (>1%) were chosen as the sample set for the second round of optimization (Table S1 of the Supporting Information). Eleven of the initially variable positions that showed very strong preference for either their native residue or one specific amino acid substitution were fixed in round 2. For the 42 remaining positions, two to seven amino acid changes were allowed. In addition, an expanded rotamer library was used to include χ angles ± one standard deviation from the most commonly observed χ angles. As in the initial round, 100 independent trajectories were performed in the second round. Overall, the results from round 2 show few significant sequence changes compared to round 1, suggesting that the expanded conformational space for side chains and the more focused library offered no clear additional benefit. Consequently, the sample set of allowed amino acid substitutions was revised for the third and final rounds of design. The structural context of individual residues and the predicted variance at each amino acid position were evaluated, and the sample set manually was refined and expanded to include homologues of high-scoring residues and native amino acids (Table S1 of the Supporting Information). Following 100 independent trajectories in round 3, sequences were analyzed on the basis of the Rosetta energy score function.

Molecular Dynamics. The predicted atom coordinates from Rosetta fixed backbone designs were scored using the CHARMM22 all atom force field.[35,36] Residues 66−71, which constitute the tail region of the protease, were omitted from our model. The structures were minimized using 1000 steps of steepest decent and conjugant gradient. The system was heated from 50 to 300 K over 30 ps, and then the molecular dynamics trajectory was recorded. Root-mean-square deviation (rmsd) calculations used an arbitrary reference at 200 ps to account for system equilibration. All calculations used the CHARMM22 force field and an implicit solvation model (SASA) and were performed using the charmm package. Simulations were run in replicate to assess reproducibility.

Gene Synthesis. The genes encoding the wild-type BPN′ propeptide sequence (WT) and the two selected designer sequences, ROS1 and ROS3, were prepared by whole-gene synthesis (DNA2.0) and shipped in the pJ201 DNA vector. For overexpression, the genes were subcloned into pET-14b (Novagen) via flanking restriction sites *Noc*I and *Spe*I. Following transformation into *Escherichia coli* DH5α, plasmid DNA from selected colonies was extracted and the sequences were verified by DNA sequencing.

Site-Directed Mutagenesis. Site-specific amino acid substitutions in the propeptide sequences were introduced by primer overlap extension mutagenesis. The synthetic WT-BPN′ gene served as a template for introducing the four mutations (A23C, K27E, V37L, and Q40C) into the disulfide-linked BPN′ propeptide variant (REF)[22] and the hydrophobic core expansion variant (T14A and K20W). Polymerase chain reaction products were cloned into pET-14b as described above and transformed into *E. coli* DH5α. Gene sequences were confirmed by DNA sequencing.

Protein Overexpression. The propeptide variants were expressed in *E. coli* BL21(DE3) pLysS or *E. coli* ArcticExpress (DE3) (Stratagene, La Jolla, CA). Briefly, individual colonies were cultured in 2 mL of LB medium containing antibiotics gentamycin (20 μg/mL) and ampicillin (100 μg/mL) overnight at 37 °C. An aliquot of overnight culture was used to inoculate 2YT medium without antibiotics, and cultures were grown at 30 °C for at least 3 h or until the OD_{600} reached 0.5−0.7. The cultures were cooled to 13 °C, and overexpression was induced by addition of IPTG to a final concentration of 1 mM. After overexpression for 24 h, cultures were centrifuged for 20 min at 4 °C and 4000*g*. The supernatant was disposed, and cell pellets were stored at −80 °C.

Protein Purification. The purification of the propeptides varies because of the difference in their isoelectric points (pI_{WT} = 9.85, pI_{ROS1} = 6.62, pI_{ROS3} = 7.9, and pI_{REF} = 9.13). In a typical purification, the cell pellet from a 500 mL culture was resuspended in 12 mL of buffer A [20 mM Tris-HCl (pH 8.5) and 50 mM NaCl for WT and 50 mM sodium phosphate (pH 6) and 50 mM NaCl for ROS1, ROS3, and REF]. To the mixture were added 100 μL of protease inhibitor cocktail (Sigma) and 10 μL of benzonase (Novagen), and the mixture was stored on ice for 15 min. Cells were lysed using sonication (eight times with 10 s pulses and 20 s pauses). After centrifugation for 30 min at 4 °C and 10000*g*, the clear lysate was further purified via two-step ion-exchange chromatography. Initially, the lysate was filtered through a HiTrap DEAE FF column (GE Healthcare). Next, the flow-through was collected

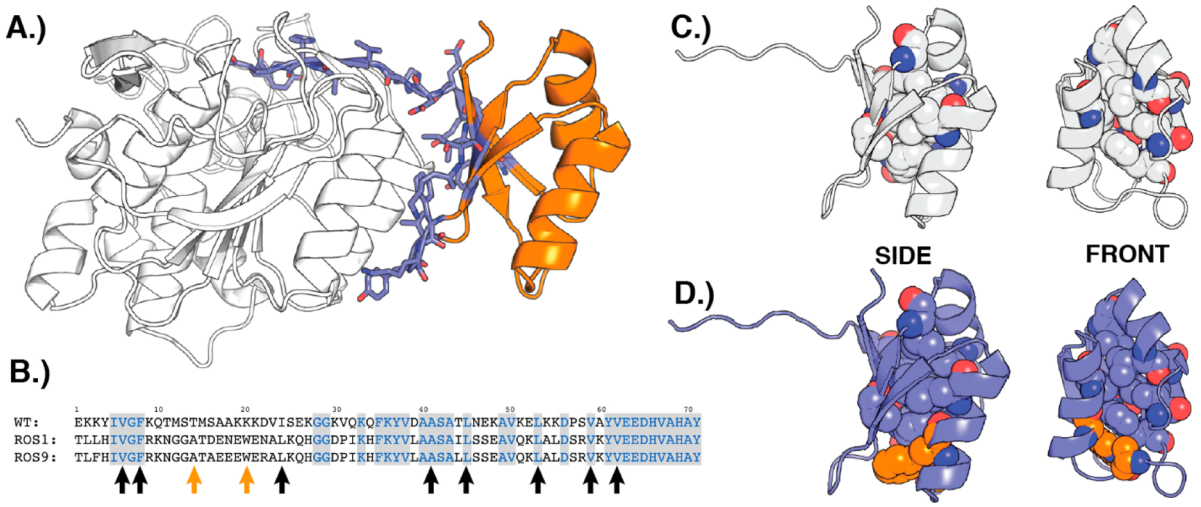


Figure 1. Protease propeptide redesign. (A) Structure of wild-type subtilisin BPN′ with bound propeptide (PDB entry 1SPB).[9] The propeptide is colored orange with residues at the propeptide−protease interface shown as purple sticks. (B) Sequence alignment of the native propeptide (WT) with Rosetta designs ROS1 and ROS3. Conserved residues are highlighted in gray, while residues of the hydrophobic core structure are marked with arrows. Positions 14 and 20 (orange arrows) represent the residues involved in the predicted core expansion. (C) Side and front views of the native propeptide with hydrophobic core residues shown in space-filling mode. (D) Computational model of ROS1 with the hydrophobic core expanded by T14A and K20W mutations (orange).

and loaded onto a HiTrap SP FF column, preequilibrated with buffer B [50 mM HEPES (pH 8) and 50 mM NaCl for WT and 50 mM sodium phosphate (pH 6) and 50 mM NaCl for ROS1, ROS3, and REF]. The column was washed with 7 column volumes (CV) of the corresponding buffer B. A linear gradient from 0 to 50% buffer C (buffer B with 1 M NaCl) over 25 CV was applied. Product fractions were combined and concentrated to ∼1 mL using a Millipore filter unit (molecular mass cutoff of 3 kDa). In a final polishing step, propeptides were purified by size exclusion chromatography {Superdex 200, 10/300 GL column, equilibrated with buffer D [50 mM Tris-HCl (pH 8) and 300 mM NaCl]; flow rate of 0.5 mL/min}. Elution of the propeptide was monitored by UV detection at 280 nm, and product fractions were combined and analyzed by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE). Propeptide samples were flash-frozen in liquid nitrogen and stored at −80 °C.

Circular Dichroism Spectroscopy. The experiments were performed using a Jasco J-810 spectropolarimeter equipped with a Peltier unit for temperature control. The path length of the cuvette used was 1 mm, with the protein concentration ranging from 25 to 40 μM as determined by the absorbance at 280 nm. All samples were prepared in 50 mM potassium phosphate buffer (pH 7−8 depending on the pI). Spectra were collected in the far-UV wavelength range at 25 °C using a scan rate of 20 nm/min, a response time of 2 s, and a bandwidth of 2 nm. The spectrum reported represents an average of three accumulation scans. For thermal denaturation experiments, the change in ellipticity at 222 nm was monitored from 4 to 80 °C with a temperature gradient of 1.0 °C/min.

Intrinsic Tryptophan Fluorescence. All measurements were taken on a FluoroMax-3 spectrophotometer (Horbia Jobin Yvon, Edison, NJ), equipped with a thermostated cell holder and a NESLAB RTE7 water bath (Thermo Electron Corp., Waltham, MA). Purified protein samples of ROS1 and ROS3 (70−140 μM) were prepared in 50 mM potassium phosphate buffer (pH 8). As a control, tryptophan (50 μM) was used in the same buffer. Samples were excited at 295 ± 0.5 nm, and their emission spectra were measured with 1 nm bandwidth resolution at wavelengths between 300 and 400 nm. Sample scans were taken in 5 °C increments from 20 to 75 °C with a 3 min sample equilibration time. All spectral data were analyzed in Origin7 to determine the maximal peak intensity and emission wavelength.

Protease Inhibition. Protease inhibition was monitored via hydrolysis of AAFP (*N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe *p*-nitroanilide) by selected proteases in the presence of various amounts of propeptide.[37] Assays were conducted with protease concentrations remaining consistent for each data set (1.8 nM for Savinase, 0.38 nM for BLAP, and 0.56 nM for Carlsberg) and 1−50-fold molar ratios of propeptide at 25 °C. Briefly, the propeptide was mixed with protease and preincubated for 0 or 20 min, followed by addition to 0.111 mM AAFP in reaction buffer [0.1 M Tris-HCl (pH 8.6) and 0.1% Brij35] and a total reaction volume of 1 mL. Upon addition of enzyme mix, substrate conversion was monitored spectrophotometrically via a change in absorbance at 410 nm over time. Substrate hydrolysis for each propeptide:protease ratio was measured in triplicate.

RESULTS AND DISCUSSION

Computational Propeptide Redesign. The Rosetta Molecular Modeling program (version 3.1) was used for redesigning the propeptide domain of subtilisin BPN′.[31] An initial search of SCOP [Structural Classification of Proteins (http://scop.mrc-lmb.cam.ac.uk/scop/)] found four crystal structures for propeptides in the subtilase family [PDB entries 1SPB (subtilisin BPN′ from *B. amyloliquefaciens*),[9] 1SCJ (subtilisin E from *Bacillus subtilis*),[12] 2Z30 (subtilisin from *Thermococcus kodakarensis*),[38] and 1T1E (pro-kumamolisin activation domain from *Bacillus* sp. MN-32)][39] and two structures of the native protease inhibitor POIA1 from *P. ostreatus* (PDB entries 1V5I and 1ITP).[19,21] On the basis of the extensive amount of prior work with propeptide from subtilisin BPN′ that sets a benchmark for our own experiments, its crystallographic coordinates when it was complexed with the

dx.doi.org/10.1021/bi300832v | *Biochemistry* 2012, 51, 8247−8255

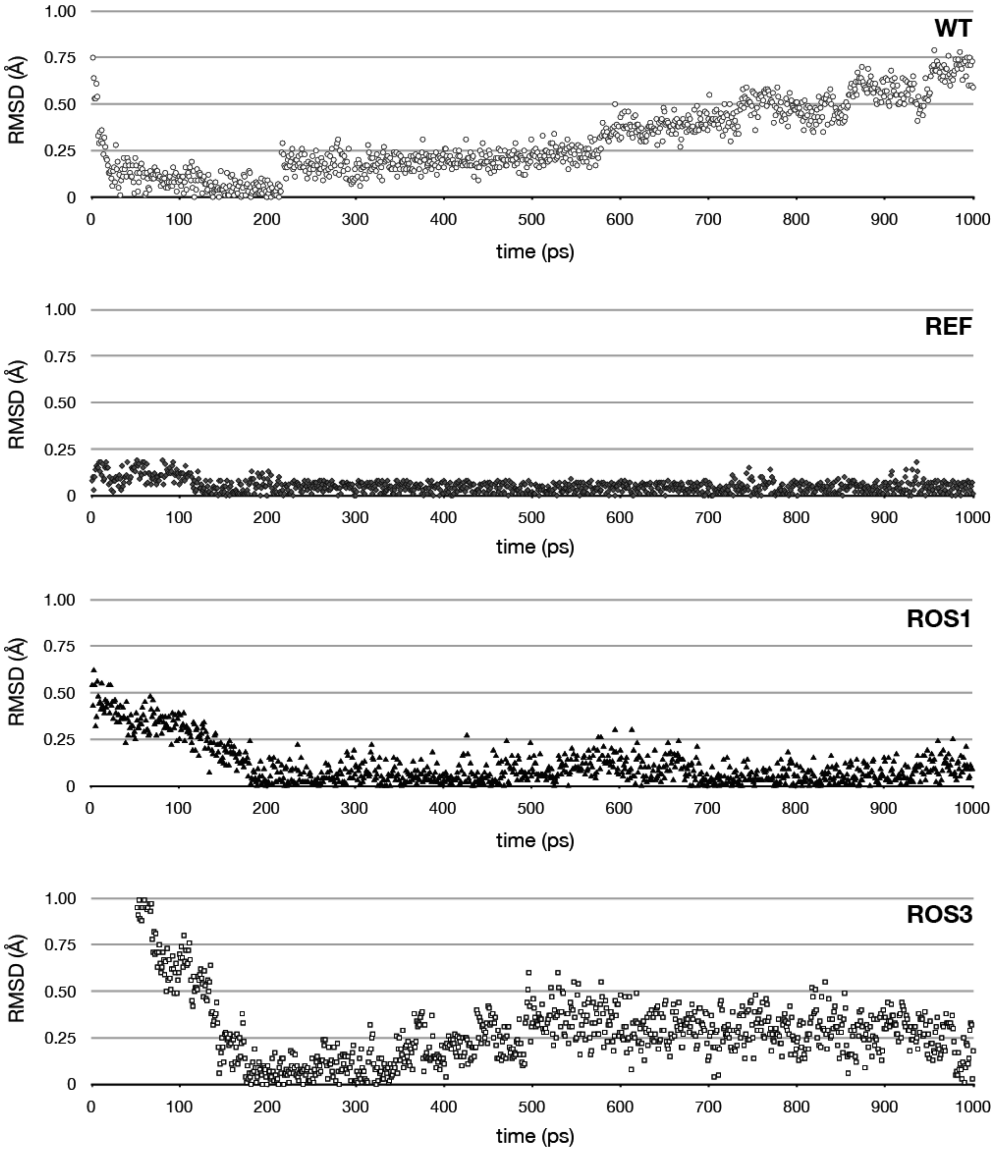


Figure 2. rmsd with respect to simulation time for the 1 ns MD simulation using wild-type BPN′ propeptide (WT) and the disulfide-linked variant (REF), as well as ROS1 and ROS3.

natural binding partner subtilisin BPN′ (PDB entry 1SPB) were selected to serve as template for three rounds of fixed backbone design (Figure 1). For the initial design, 54 noninterface residues of the 71-residue propeptide sequence were allowed to sample the side chains of all natural amino acids (Table S1 of the Supporting Information). To prevent perturbation of the propeptide−protease interface, changes in the 17 remaining residues that directly contact the protease core structure were limited to side chain rotamers of the native amino acids. Over three rounds of Rosetta design, the propeptide sequence remained 40−50% identical to that of the parental BPN′ propeptide. Beyond the 17 originally conserved residues, an additional 12−18 positions showed a strong preference for the native amino acid while positions 36−42 experienced amino acid substitutions (Table S1 of the Supporting Information). Those amino acid changes ranged from relatively conservative substitutions (i.e., K9R and I24L) to complete charge reversion (i.e., K19E, K21E, and D22R), as well as functional shifts (i.e., K3F, M12G, S13G, and K20W). Nevertheless, no dramatic changes with regard to overall favorable net surface charges, hydrogen bonding networks, and hydrophobicity nor correlation between Rosetta's score function and amino acid substitution patterns could be observed among third-round sequences. The exception is an intriguing design feature that emerged in the extended loop region between β1 and α1, as well as the lower portion of helix 1 (residues 9−20) near the N-terminus of the propeptide variants. This region shows elevated *B* factors in the crystal structure of the WT propeptide, suggesting increased conformational flexibility that could at least in part be responsible for the structural instability of the native domain. In all third-round variants, Rosetta replaced the -MSTxxxxKKK- region (residues 12−21) with -GGAxxxx(E/K)W(E/K)-. A review of the corresponding model structure indicates that the glycines provide the necessary conformational flexibility for a tight turn in the protein structure, while substitutions T14A and K20W appear to translate into an expansion of the hydrophobic core of the propeptide (Figure 1C,D). The resulting structural integration of the extended loop region could be beneficial to overall protein stability. Moreover, the strategy to stabilize the propeptide through an

expansion of the hydrophobic core is clearly distinct from that of the previously reported covalently linked disulfide variant, offering in principle the opportunity for tunability of protein stability through modifications of noncovalent interactions in proteins, including hydrophobic core and/or surface residues.

Separately, we validated our Rosetta propeptide designs by assessing the structural integrity of propeptide variants via molecular dynamics (MD) simulation using a CHARMM22 force field. Rather than performing such simulations on all designs, we selected representative sequences on the basis of a phylogenetic analysis of third-round sequences using ClustalW2.[40] The analysis reveals a grouping of the propeptide sequences into eight clusters (Figure S1 of the Supporting Information). The differences in sequence among the eight clusters are relatively small and limited to various combinations of amino acid changes at 10 positions. In addition to wild-type BPN′ propeptide (WT) and the disulfide-linked variant of Bryan and co-workers[22] (REF), a representative from each cluster was picked [ROS1−8 (Figure S2 of the Supporting Information)] and used for the MD simulations. The predicted difference in protein stability of the propeptide variants translates into distinct, time-dependent trajectories for their rmsd after an arbitrary burn-in period of 200 ps to reach the equilibrated structure (Figure 2 and Figure S3 of the Supporting Information). As shown in Figure 2, the WT sequence does not appear to be stable, atrophying to the disordered structure. In contrast, the REF sequence seems to be very stable, as deviant conformers quickly recover to the equilibrated state. The artificial disulfide bond, covalently stitching critical secondary structures, may confer this increased stability. The MD data for the eight Rosetta designs fall between these two extremes, showing mostly favorable, rebounding trajectories. These results are consistent with the anticipated improvement in the structural stability of variants compared to WT propeptide yet reflect greater conformational variance than those for REF (Figure 2 and Figure S3 of the Supporting Information). The only exception is ROS4, for which simulation suggests a similar progression toward a disorderly state as seen for WT. On the basis of the computational predictions, we chose two Rosetta designs, the top-scoring ROS1 and the "second-tier candidate" ROS3, for in-depth experimental evaluation.

Expression and Purification of Designer Propeptides. The gene sequences corresponding to the predicted top-scoring propeptide designs ROS1 and ROS3 as well as WT propeptide were prepared by whole-gene DNA synthesis. All nucleotide sequences were codon-optimized for protein expression in *E. coli*. Separately, we generated the disulfide-linked propeptide variant that served as a positive control in our study. The four required mutations (A23C, K27E, V37L, and Q40C) were introduced into the WT sequence, yielding the reference propeptide (REF). Soluble expression of all four peptide sequences was readily accomplished in *E. coli* ArcticExpress (DE3) (for ROS1 and ROS3) or *E. coli* BL21(DE3) pLysS (for WT and REF), yielding protein at a concentration of ∼5 mg/L of culture for ROS1 and ROS3 and 15−20 mg/L of culture for WT and REF. The difference in soluble protein expression levels for the Rosetta designs compared to the WT and REF sequences is significant and interesting in light of the notably improved thermostability of the former (see below). A possible explanation for the phenomenon could be Rosetta's focus on the final, folded propeptide structure without consideration of the actual folding pathway. Nevertheless, soluble protein could be purified by a combination of ion-exchange and gel filtration chromatography, yielding the individual propeptides at >95% purity based on SDS−PAGE analysis. The gel filtration experiments also confirmed previous reports of a propeptide monomer−dimer equilibrium, indicating a 70:30 ratio in favor of monomeric ROS1, ROS3, and REF.

Intrigued by the idea of stabilizing the propeptide through the expansion of the hydrophobic core region, we were also curious about whether extensive mutagenesis at more than 50% of the variable positions as suggested by Rosetta was indeed necessary for enhanced protein stability. Alternatively, substantive improvements might be possible through fewer, more specific substitutions in the extended loop region. We thus decided to build a WT variant with only the two site-specific amino acid substitutions at positions 14 (T14A) and 20 (K20W). In the absence of obvious steric problems to accommodate these changes, the two substitutions were introduced by site-directed mutagenesis of the WT gene sequence. Expression of the corresponding propeptide in *E. coli* ArcticExpress (DE3) was attempted at various temperatures and under various induction conditions yet failed to produce any detectable amounts of soluble propeptide. Our results demonstrate that the simple double mutant WT propeptide with the core expansion substitutions alone is insufficient to accomplish the anticipated propeptide stabilization and emphasizes the benefits of additional, compensatory substitutions necessary for expression of soluble protein.

Structural Characterization of Propeptides. The secondary structure content of the WT propeptide as well as ROS1, ROS3, and REF was initially investigated by far-UV circular dichroism (CD) spectroscopy (Figure 3). Spectral scans were recorded from 190 to 260 nm at ambient temperature. In line with previous findings, the WT protein spectrum was consistent with random coil and is assumed to exist in a largely disordered state. In contrast, the spectra for ROS1, ROS3, and REF were consistent with mixed α-helix/β-sheet structures as predicted on the basis of the secondary

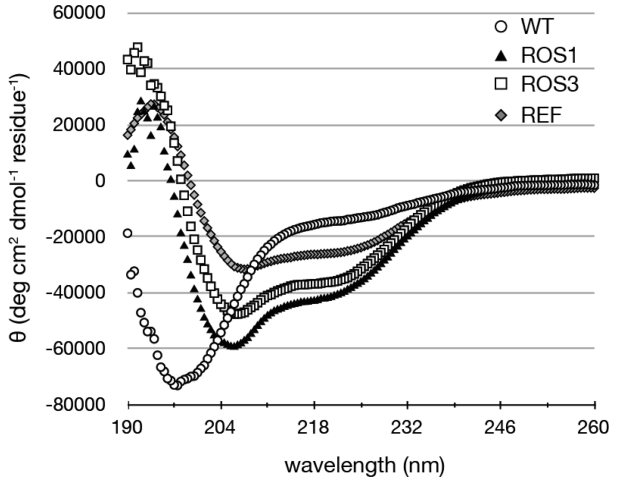


Figure 3. Far-UV circular dichroism spectra (190−260 nm) of wild-type BPN′ (WT), Rosetta design ROS1 and ROS3, and the disulfide-linked variant (REF). While the WT sample resembles a random coil, the spectra of ROS1, ROS3, and REF exhibit clear signatures of mixed α-helix/β-sheet secondary structures. Spectra were recorded on protein samples (25−40 μM) in potassium phosphate buffer (pH 7−8) at ambient temperature.

dx.doi.org/10.1021/bi300832v | *Biochemistry* 2012, 51, 8247−8255

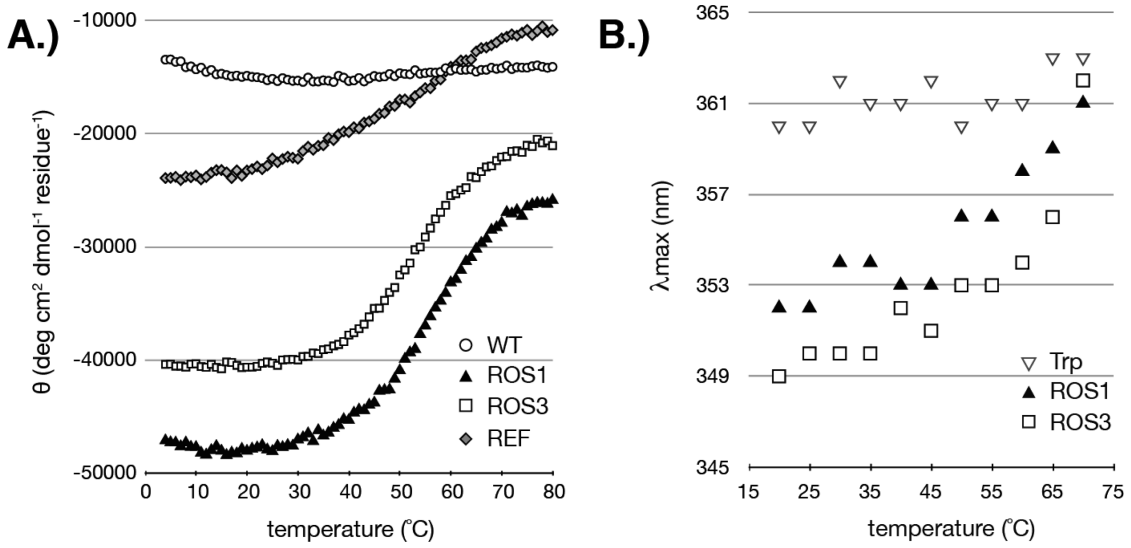


Figure 4. Irreversible thermal denaturation curves of wild-type BPN′ (WT), Rosetta designs ROS1 and ROS3, and the disulfide-linked variant (REF). (A) Thermal unfolding detected at 222 nm by CD spectroscopy from 4 to 80 °C. While WT is unfolded at all temperatures, data for ROS1, ROS3, and REF were analyzed by curve fitting to a two-state model using the Boltzman equation in Origin7 (OriginLab, Northampton, MA). Spectra were recorded on protein samples (25−40 μM) in potassium phosphate buffer (pH 7−8). (B) Thermal unfolding detected by intrinsic tryptophan fluorescence via the emission maximal shift of the unique W20 residue in ROS1 and ROS3. Spectra were recorded on protein samples (70−140 μM) in potassium phosphate buffer (pH 8) in 5 °C increments at the indicated temperatures. Free Trp (50 μM) in the same buffer served as a control.

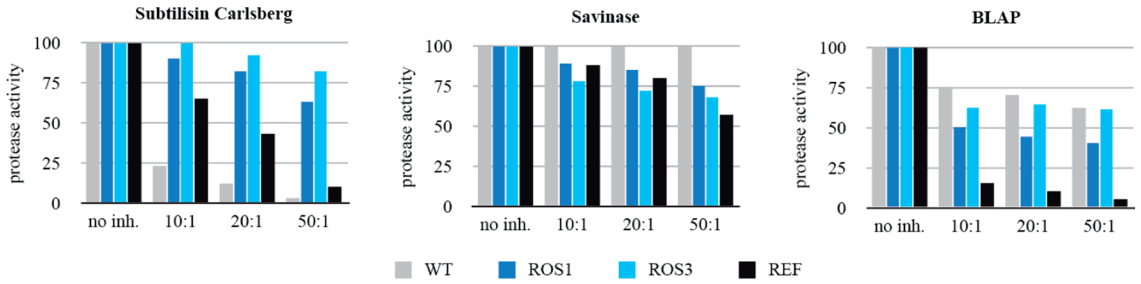


Figure 5. Comparison of inhibitory activity for ROS1 and ROS3 with WT and REF against three subtilisin proteases (Carlsberg, Savinase, and BLAP). The standard error is less that 20% for all inhibitory activities. Residual protease activity was determined by the following equation: activity = [(activity in the presence of propeptide)/(activity in the absence of propeptide)] × 100%.

structure elements observed in the BPN′ propeptide crystal structure as part of the pro-protease structure (PDB entry 1SPB).[9]

Next the stability of all four propeptides was assessed by thermodenaturation (Figure 4A). Changes in secondary structure were monitored at 222 nm from 4 to 80 °C in the CD spectrophotometer. As predicted from the far-UV CD spectrum, WT propeptide was unfolded at all temperatures with no significant melting transition while the three remaining protein variants showed temperature-induced denaturation profiles. The broad profile for REF indicates a complex unfolding pattern with a denaturation midpoint (T_M) of ~47 °C. In comparison, the unfolding transition for ROS1 and ROS3 was notably sharper and shifted to higher T_M values of 55 and 53 °C, respectively. The experimentally determined trend in protein stability for the Rosetta designs is consistent with the computational predictions. In addition to the CD thermodenaturation experiments, we also studied tertiary structure changes in ROS1 and ROS3 at elevated temperatures via intrinsic tryptophan fluorescence (Figure 4B). While WT and REF lack any tryptophan residue in their sequences, ROS1 and ROS3 carry a unique tryptophan residue at position 20. On the basis of our models, the residue is positioned at the edge of the propeptide hydrophobic core and should display an environment-dependent red shift of its emission spectrum upon protein unfolding and increased solvent exposure. These predictions were confirmed experimentally for both propeptide variants (Figure 4B). At ambient temperature, ROS1 and ROS3 had tryptophan emission maxima of 352 and 349 nm, respectively. With a gradual increase in temperature, their emission maxima shifted exponentially toward higher wavelengths to reach ~362 nm at 70 °C. For comparison, the emission maximum of free tryptophan in the same buffer solution was near 360 nm at ambient temperature and showed only a small temperature-induced red shift to 362 nm at 70 °C. If we assume that both ROS1 and ROS3 are largely unfolded at 70 °C, our fluorescence data give estimated T_M values between 55 and 60 °C for the two proteins, consistent with the CD data for the loss of secondary structure at elevated temperatures. Overall, the CD and fluorescence data support the conclusion that the redesigned propeptide sequences ROS1 and ROS3 are stable and structurally well-defined proteins.

Functional Characterization of Propeptides. In parallel to the structural characterization of our designer propeptides,

dx.doi.org/10.1021/bi300832v | *Biochemistry* 2012, 51, 8247−8255

we evaluated the inhibitory activity of ROS1 and ROS3, as well as WT and REF against three serine proteases, subtilisin Carlsberg, Savinase, and BLAP. Crystal structures among protease complexes have <0.4 Å interface rmsd, and sequence similarity in this region maintains complementary patches of buried hydrophobic and charged interactions, suggesting similar driving forces among subtilisin family members. Subtilisin BPN′ was not included in our study, even though it served as a template for our redesign, as the protease is no longer commercially available. The performance of the propeptide variants was tested in a standard spectrophotometric AAPF activity assay, monitoring the proteolytic activity of individual hydrolases in the presence of different propeptide inhibitor molar ratios.[37]

As shown in Figure 5, all three proteases were inhibited by WT, ROS1, ROS3 and REF albeit to various degrees. The one exception was WT propeptide which did not affect Savinase activity. Against subtilisin Carlsberg, the WT propeptide proved an effective inhibitor with less than 15% proteolytic activity at 20:1 inhibitor:protease ratio. The result was only matched by REF at the highest ratio (50:1) tested in our assay. In contrast, ROS1 and ROS3 show only slight inhibitor activity (60−80% residual proteolytic activity at 50:1). For BLAP, the pattern of inhibitor performance changes significantly. The presence of REF lowers protease activity to <20% at a 10:1 ratio while ROS1 reduces AAPF hydrolysis to ∼50% at 10:1 and ∼40% at 50:1 inhibitor:protease. ROS3 and WT are less effective, only reaching 30−40% inhibition at a 50:1 ratio. In the case of Savinase, the two Rosetta designs and REF perform similarly, suppressing protease activity to ∼70% at the 50:1 ratio. Despite of their moderate inhibitory performance, the results for the three propeptide variants are relevant in light of no detectable inhibitory activity with WT.

Overall, the functional study demonstrates that all four propeptide variants can inhibit selected proteases. Nevertheless, there is no clear correlation between improved thermostability of the propeptides and their effectiveness as inhibitors in our data set. Beyond measuring the inhibitory function of the propeptide variants immediately after they had been mixed with protease, we also explored the time-dependent change in inhibitor activity. Preincubation of individual propeptide variants with proteases for 20 min prior to substrate addition might reveal a stronger correlation between propeptide stability and inhibitor function as the increased structural integrity of the propeptide can slow its degradation by the protease, in turn prolonging inhibitor activity. Subsequent AAPF hydrolysis measurements of the preincubated protein mixtures did detect a reduction in inhibitor activity among all tested propeptide, yet changes were generally within the margin of error (data not shown). Additional experiments exploring longer preincubation times for each protease−inhibitor pair might be necessary for detecting significant differences in inhibitor function in correlation to propeptide stability.

■ CONCLUSIONS

Our study has demonstrated the successful application of computational protein design in combination with MD simulations in generating thermostable variants of the subtilisin BPN′ propeptide domain. The experimental evaluation of the structural properties of ROS1 and ROS3 is consistent with relative rankings based on the arbitrary energy scores in RosettaDesign. Although the data set will benefit from the experimental characterization of additional propeptide designs, our results are consistent with the idea that in silico predictions correlate with propeptide stability. Despite substantial improvements in the structural integrity as reflected by the increased thermostability, the protease inhibition studies nevertheless highlight the current discrepancies between structure and function. More specifically, the original hypothesis in the literature that stabilization of the propeptide will translate into more potent protease inhibitors is clearly not supported by our results. Even for WT and REF, which share high degrees of sequence and structural identity, the inhibitory activity for different proteases varies significantly with no recognizable correlation to the elevated stability of REF. Comparison of the crystal structures for a disulfide-linked propeptide domain closely related to REF (PDB entry 3CNQ)[41] and the WT structure (PDB entry 1SPB)[9] shows an rmsd of 0.94 Å over the entire length of the 71-residue propeptide, yet despite these striking structural similarities, the inhibitory effects of these two propeptides on the proteases in our experiments are notably different. We attribute these uncertainties to discrepancies related to the origin of the propeptide model (subtilisin BPN′) and the tested proteases, as well as small differences in structure and dynamics not accurately captured in the crystallographic information and the computational model. Improved future designs might be possible with the help of advanced Rosetta models considering the flexible backbone and incorporation of interactions at the protease−propeptide interface. Separately, high-resolution nuclear magnetic resonance (NMR) studies will be insightful for validating the accuracy of the Rosetta designs and investigating possible conformational differences between the native propeptide domain and the redesigned structures. Preliminary NMR experiments with the two designer propeptides do indicate a defined secondary structure similar to that of REF (data not shown). Finally, we postulate that the C-terminal portion of the propeptide that was left unchanged for all variants in our study might play an important role in the context of the inhibitory function of these proteins. The region's compatibility with and affinity for the substrate recognition cleft flanking the active site of individual proteases as well as conformational differences between these propeptides and native protease inhibitors such as POIA1 will be the subject of future studies.

■ ASSOCIATED CONTENT

Supporting Information

Details on the computation design and analysis of propeptide variants. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sal2@emory.edu. Telephone: (404) 712-2170.

Funding

This work was in part supported by funds from the National Science Foundation Center for Pharmaceutical Manufacturing and Formulation (I/UCRC 0832469) and through a collaboration with Henkel AG & Co. KGaA (Duesseldorf, Germany).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Bordusa, F. (2002) Proteases in organic synthesis. *Chem. Rev. 102*, 4817−4868.

(2) Savile, C. K., Magloire, V. P., and Kazlauskas, R. J. (2005) Subtilisin-catalyzed resolution of N-acyl arylsulfinamides. *J. Am. Chem. Soc. 127*, 2104−2113.

(3) Gallagher, T., Ruan, B., London, M., Bryan, M. A., and Bryan, P. N. (2009) Structure of a switchable subtilisin complexed with a substrate and with the activator azide. *Biochemistry 48*, 10389−10394.

(4) Maurer, K. H. (2004) Detergent proteases. *Curr. Opin. Biotechnol. 15*, 330−334.

(5) Stoner, M. R., Dale, D. A., Gualfetti, P. J., Becker, T., Manning, M. C., Carpenter, J. F., and Randolph, T. W. (2004) Protease autolysis in heavy-duty liquid detergent formulations: Effects of thermodynamic stabilizers and protease inhibitors. *Enzyme Microb. Technol. 34*, 114−125.

(6) Lee, S. L., Alexander, R. S., Smallwood, A., Trievel, R., Mersinger, L., Weber, P. C., and Kettner, C. (1997) New inhibitors of thrombin and other trypsin-like proteases: Hydrogen bonding of an aromatic cyano group with a backbone amide of the P1 binding site replaces binding of a basic side chain. *Biochemistry 36*, 13180−13186.

(7) Smoum, R., Rubinstein, A., Dembitsky, V. M., and Srebnik, M. (2012) Boron Containing Compounds as Protease Inhibitors. *Chem. Rev. 112* (7), 4156−4220.

(8) Ganz, P. J., Bauer, M. D., Sun, Y. P., Fieno, A. M., Grant, R. A., Correa, P. E., Laskowski, M., and Saunders, C. W. (2004) Stabilized variant of *Streptomyces* subtilisin inhibitor and its use in stabilizing subtilisin BPN′. *Protein Eng., Des. Sel. 17*, 333−339.

(9) Gallagher, T., Gilliland, G., Wang, L., and Bryan, P. (1995) The Prosegment-Subtilisin Bpn Complex: Crystal-Structure of a Specific Foldase. *Structure 3*, 907−914.

(10) Eder, J., Rheinnecker, M., and Fersht, A. R. (1993) Folding of Subtilisin Bpn′: Role of the Pro-Sequence. *J. Mol. Biol. 233*, 293−304.

(11) Shinde, U., Li, Y. Y., Chatterjee, S., and Inouye, M. (1993) Folding Pathway Mediated by an Intramolecular Chaperone. *Proc. Natl. Acad. Sci. U.S.A. 90*, 6924−6928.

(12) Jain, S. C., Shinde, U., Li, Y. Y., Inouye, M., and Berman, H. M. (1998) The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 angstrom resolution. *J. Mol. Biol. 284*, 137−144.

(13) Strausberg, S., Alexander, P., Wang, L., Schwarz, F., and Bryan, P. (1993) Catalysis of a Protein-Folding Reaction: Thermodynamic and Kinetic-Analysis of Subtilisin BPN′ Interactions with Its Propeptide Fragment. *Biochemistry 32*, 8112−8119.

(14) Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G., and Gallagher, T. (1995) Catalyis of a Protein-Folding Reaction: Mechanistic Implications of the 2.0 Å Structure of the Subtilisin-Prodomain Complex. *Biochemistry 34*, 10310−10318.

(15) Kojima, S., Iwahara, A., and Yanai, H. (2005) Inhibitor-assisted refolding of protease: A protease inhibitor as an intramolecular chaperone. *FEBS Lett. 579*, 4430−4436.

(16) Ohta, Y., Hojo, H., Aimoto, S., Kobayashi, T., Zhu, X., Jordan, F., and Inouye, M. (1991) Pro-Peptide as an Intermolecular Chaperone: Renaturation of Denatured Subtilisin-E with a Synthetic Pro-Peptide. *Mol. Microbiol. 5*, 1507−1510.

(17) Kojima, S., Yanai, H., and Miura, K. (2001) Accelerated refolding of subtilisin BPN′ by tertiary-structure-forming mutants of its propeptide. *J. Biochem. 130*, 471−474.

(18) Kojima, S., Minagawa, T., and Miura, K. (1997) The propeptide of subtilisin BPN′ as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity. *FEBS Lett. 411*, 128−132.

(19) Sasakawa, H., Yoshinaga, S., Kojima, S., and Tamura, A. (2002) Structure of POIA1, a homologous protein to the propeptide of subtilisin: Implication for protein foldability and the function as an intramolecular chaperone. *J. Mol. Biol. 317*, 159−167.

(20) Kojima, S., Deguchi, M., and Miura, K. (1999) Involvement of the C-terminal region of yeast proteinase B inhibitor 2 in its inhibitory action. *J. Mol. Biol. 286*, 775−785.

(21) Lee, W. C., Kikkawa, M., Kojima, S., Miura, K., and Tanokura, M. (2005) Crystal structure of serine protease inhibitor POIA1 in complex with subtilisin BPN′. Data deposited in PDB only.

(22) Ruan, B., Hoskins, J., Wang, L., and Bryan, P. N. (1998) Stabilizing the subtilisin BPN′ pro-domain by phage display selection: How restrictive is the amino acid code for maximum protein stability? *Protein Sci. 7*, 2345−2353.

(23) Wang, L., Ruan, B., Ruvinov, S., and Bryan, P. N. (1998) Engineering the independent folding of the subtilisin BPN′ pro-domain: Correlation of pro-domain stability with the rate of subtilisin folding. *Biochemistry 37*, 3165−3171.

(24) Ruan, B., Hoskins, J., and Bryan, P. N. (1999) Rapid folding of calcium-free subtilisin by a stabilized pro-domain mutant. *Biochemistry 38*, 8562−8571.

(25) Ruvinov, S., Wang, L., Ruan, B., Almog, O., Gilliland, G. L., Eisenstein, E., and Bryan, P. N. (1997) Engineering the independent folding of the subtilisin BPN′ prodomain: Analysis of two-state folding versus protein stability. *Biochemistry 36*, 10414−10421.

(26) Kojima, S., Minagawa, T., and Miura, K. (1998) Tertiary structure formation in the propeptide of subtilisin BPN′ by successive amino acid replacements and its close relation to function. *J. Mol. Biol. 277*, 1007−1013.

(27) Korkegian, A., Black, M. E., Baker, D., and Stoddard, B. L. (2005) Computational thermostabilization of an enzyme. *Science 308*, 857−860.

(28) Shah, P. S., Hom, G. K., Ross, S. A., Lassila, J. K., Crowhurst, K. A., and Mayo, S. L. (2007) Full-sequence computational design and solution structure of a thermostable protein variant. *J. Mol. Biol. 372*, 1−6.

(29) Dantas, G., Kuhlman, B., Callender, D., Wong, M., and Baker, D. (2003) A large scale test of computational protein design: Folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol. 332*, 449−460.

(30) Dantas, G., Corrent, C., Reichow, S. L., Havranek, J. J., Eletr, Z. M., Isern, N. G., Kuhlman, B., Varani, G., Merritt, E. A., and Baker, D. (2007) High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *J. Mol. Biol. 366*, 1209−1221.

(31) Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popovic, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzymol. 487*, 545−574.

(32) Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364−1368.

(33) Dunbrack, R. L., Jr., and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci. 6*, 1661−1681.

(34) Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R., and Ben-Tal, N. (2004) ConSeq: The identification of functionally and structurally important residues in protein sequences. *Bioinformatics 20*, 1322−1324.

(35) Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) Charmm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem. 4*, 187−217.

(36) Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem. 30*, 1545−1614.

(37) DelMar, E. G., Largman, C., Brodrick, J. W., and Geokas, M. C. (1979) A sensitive new substrate for chymotrypsin. *Anal. Biochem. 99*, 316−320.
(38) Tanaka, S., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. (2007) Four new crystal structures of Tk-subtilisin in unautoprocessed, autoprocessed and mature forms: Insight into structural changes during maturation. *J. Mol. Biol. 372*, 1055−1069.
(39) Comellas-Bigler, M., Maskos, K., Huber, R., Oyama, H., Oda, K., and Bode, W. (2004) 1.2 Å crystal structure of the serine carboxyl proteinase pro-kumamolisin; structure of an intact pro-subtilase. *Structure 12*, 1313−1323.
(40) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics 23*, 2947−2948.
(41) Ruan, B., London, V., Fisher, K. E., Gallagher, D. T., and Bryan, P. N. (2008) Engineering substrate preference in subtilisin: Structural and kinetic analysis of a specificity mutant. *Biochemistry 47*, 6628−6636.