JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

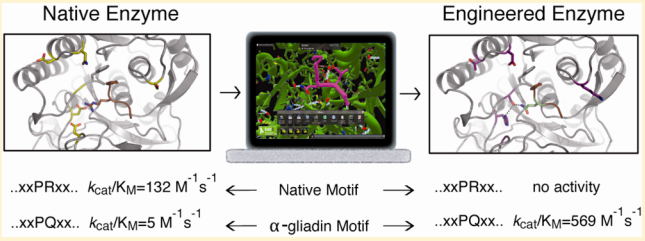
Computational Design of an α-Gliadin Peptidase

Sydney R. Gordon,[†,‡] Elizabeth J. Stanley,[†,‡] Sarah Wolf,[†] Angus Toland,[†] Sean J. Wu,[†] Daniel Hadidi,[†] Jeremy H. Mills,[†] David Baker,[†,#] Ingrid Swanson Pultz,*[,§] and Justin B. Siegel*[,†,⊥,∥,▽]

[†]Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States
[#]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, United States
[§]Department of Microbiology, University of Washington, Seattle, Washington 98195, United States
[⊥]Department of Biochemistry & Molecular Medicine, University of California, Davis, California, United States
[∥]Department of Chemistry, University of California, Davis, California, United States
[▽]Genome Center, University of California, Davis, California, United States

Supporting Information

ABSTRACT: The ability to rationally modify enzymes to perform novel chemical transformations is essential for the rapid production of next-generation protein therapeutics. Here we describe the use of chemical principles to identify a naturally occurring acid-active peptidase, and the subsequent use of computational protein design tools to reengineer its specificity toward immunogenic elements found in gluten that are the proposed cause of celiac disease. The engineered enzyme exhibits a k_{cat}/K_M of 568 M^{-1} s^{-1}, representing a 116-fold greater proteolytic activity for a model gluten tetrapeptide than the native template enzyme, as well as an over 800-fold switch in substrate specificity toward immunogenic portions of gluten peptides. The computationally engineered enzyme is resistant to proteolysis by digestive proteases and degrades over 95% of an immunogenic peptide implicated in celiac disease in under an hour. Thus, through identification of a natural enzyme with the pre-existing qualities relevant to an ultimate goal and redefinition of its substrate specificity using computational modeling, we were able to generate an enzyme with potential as a therapeutic for celiac disease.



INTRODUCTION

The application of computational protein design tools has been demonstrated to introduce functional properties beyond those obtained by natural evolution, such as producing enzymes that perform functions not found in nature, altered specificity of proteins for their binding partners, and the *de novo* design of fold topologies.[1−3] Given the success of these early engineering efforts, the application of protein design tools holds great promise for next-generation protein therapeutics. Celiac disease is a condition that provides an ideal model for enzyme therapeutic development through computational protein design; the illness sets forth clear chemical and catalytic demands of a desired therapeutic, which can be built into an enzyme template through rational design.

Celiac disease is characterized by an inflammatory reaction in the digestive tract to α-gliadin, an important component of the glycoprotein gluten, which is found in any food containing wheat, rye, or barely.[4,5] Upon ingestion, α-gliadin is partially degraded by digestive proteases to oligopeptides that are resistant to further proteolysis due to their unusually high proline and glutamine content[5] (Figure 1). These proteolytically resistant oligopeptides that are highly enriched in a proline−glutamine (PQ) motif are believed to trigger an autoimmune response, which elicits many of the symptoms in celiac patients.[6] An ideal oral enzyme therapeutic (OET) for celiac disease would have the following traits: (1) optimal activity at the pH of the stomach after a meal (in the range of 2−4);[7] (2) resistance to common digestive proteases; (3) facile recombinant production in a soluble form; and (4) specificity for the common proline−glutamine (PQ) motif found in immunogenic α-gliadin oligopeptides.[6,8] While there are several OETs being explored for the treatment of celiac disease,[6,9−11] none conform to all of these properties.

Here we report the use of computational protein design to engineer an endopeptidase with all of the desired traits for an OET for celiac disease. Using knowledge of the chemical principles required of an OET for celiac disease (e.g., high activity at low pH), we identified an endopeptidase that is highly active in acidic conditions: kumamolisin-As (KumaWT), from the acidophilic bacterium *Alicyclobacillus sendaiensis*. We then used the Rosetta Molecular Modeling Suite to identify a subset of mutations likely to enhance its activity for the desired oligopeptide specificity. The computationally designed enzyme (designated KumaMax) exhibited a k_{cat}/K_M of 568 M^{-1} s^{-1}, which represents a 116-fold increased proteolytic activity and

Received: September 24, 2012
Published: November 15, 2012

   dx.doi.org/10.1021/ja3094795 | J. Am. Chem. Soc. 2012, 134, 20513−20520

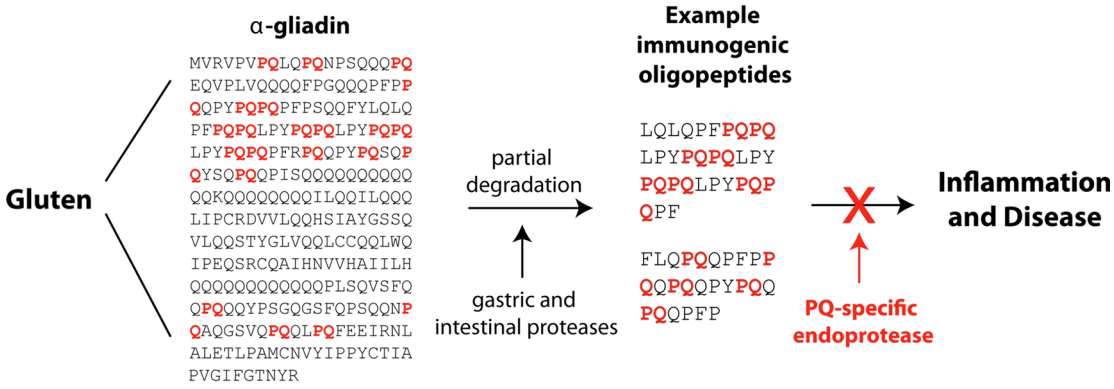


Figure 1. Schematic depicting the role of enzyme therapeutics in the treatment of celiac disease. Gluten is comprised of many glycoproteins including α-gliadin. Partial proteolysis of α-gliadin results in protease-resistant peptides enriched in a PQ dipeptide motif that can lead to inflammation and disease. An enzyme that is functional in the stomach and capable of specifically degrading the immunogenic peptides could potentially act as a therapeutic for this disease.

an 877-fold switch in substrate specificity for the targeted PQ motif compared with KumaWT. In addition, both KumaWT and KumaMax were produced at approximately 30 mg/L of culture in *Escherichia coli* without the need for refolding. Finally, we demonstrate that KumaMax is resistant to common gastric proteases. These combined properties make the engineered peptidase a promising candidate as an oral therapeutic for celiac disease.

RESULTS

Identification and Computational Design of an α-Gliadin Endopeptidase. To engineer a peptidase that can degrade gliadin peptides under gastric conditions, we first focused on identifying a suitable naturally occurring endopeptidase that could be used as a starting point for our engineering efforts. Ideally, such an enzyme would exhibit high stability and activity at an acidic pH, with specificity for a dipeptide amino acid motif similar to the PQ motif found in the immunogenic gluten peptides (Figure 1). We searched the MEROPS Peptide Database to identify enzyme families that are known to be active in acidic pH. These are the aspartic and glutamic peptidases, and the S10 and S53 families of serine peptidases. We further narrowed our search by selecting endopeptidases, produced in a soluble form using an *E. coli* platform, that have an available crystal structure in complex with their peptide substrates. Finally, we examined the amino acid specificities of the candidate endopeptidases to identify those that are specific to an amino acid motif similar to our targeted PQ motif.

On the basis of the above criteria, we identified the enzyme kumamolisin-As (KumaWT) as an excellent candidate. KumaWT is a serine endopeptidase that demonstrates a serine−glutamate−aspartate catalytic triad instead of the serine−histidine−aspartate triad of traditional serine proteases. The involvement of a glutamate (pK_a of 4.1), instead of a histidine (pK_a of 6.5), in the catalytic triad allows this enzyme to function at low pH. Indeed, KumaWT exhibits optimal activity over the pH range of 2−4,[12] which matches the approximate pH range of the human stomach after a meal.[7] KumaWT also demonstrates high stability and activity at the physiologically relevant temperature of 37 °C.[13] The purification of this enzyme yields significant quantities of soluble protein using standard recombinant protein production methods in *E. coli*,[13] an important property both for screening mutant libraries and for its eventual production for use as an OET. Finally, KumaWT naturally recognizes a specific dipeptide motif as opposed to a single amino acid.[13] This property is potentially important for an OET meant to be taken during digestion, since dipeptide specificity should result in reduced competitive inhibition from other food-derived peptides in the stomach or off-target effects due to degradation of other necessary proteins in the gut.

An effective OET for celiac disease would ideally be specific for a PQ motif, due to the frequent occurrence of this dipeptide in immunogenic gluten-derived α-gliadin oligopeptides (Figure 1). KumaWT exhibits specificity for proline at the P2 position of its peptide substrate, matching the P2 residue of interest for the degradation of immunogenic α-gliadin peptides.[13] In the P1 site, KumaWT prefers the positively charged amino acids arginine or lysine.[13] Despite this preference, KumaWT is also capable of recognizing glutamine at the P1 position, albeit at a significantly decreased level compared with its recognition of arginine or lysine.[13] This low level of activity for glutamine at the P1 position suggests that KumaWT might be amenable to re-engineering to prefer glutamine at this position. At the P1′ site, KumaWT demonstrates broad specificity, which is desirable since the residue in the position after the PQ motif varies among the different immunogenic peptides (Figure 1).[14]

Given these characteristics of KumaWT, we sought to computationally redesign the S1 binding pocket of KumaWT such that it would prefer a PQ dipeptide motif over the native PR or PK substrates. Using the Rosetta Molecular Modeling Suite,[15] we modeled the PR dipeptide in the S1 binding pocket of the KumaWT crystal structure (PDB-ID 1T1E). This revealed that two negatively charged amino acids, D358 and D368, likely facilitate binding of the positively charged amino acids in the P1 position (Figure 2A). The native specificity for proline at P2 appears to be derived in large part from a hydrophobic interaction of this residue with the aromatic ring of W318 in the S2 pocket of the enzyme. Because specificity at the P2 position for proline is desired for OET, we maintained this native tryptophan during the design of the S1 pocket.

To redesign the substrate specificity of the S1 pocket to prefer glutamine at the P1 position, we modeled mutations in the KumaWT binding pocket using the Foldit interface to the Rosetta Molecular Modeling Suite.[16] A tetrapeptide that represents a common immunogenic motif found throughout α-gliadin, PQLP, was modeled into the P2 to P2′ active site positions. The crystal structure already contained a polypeptide

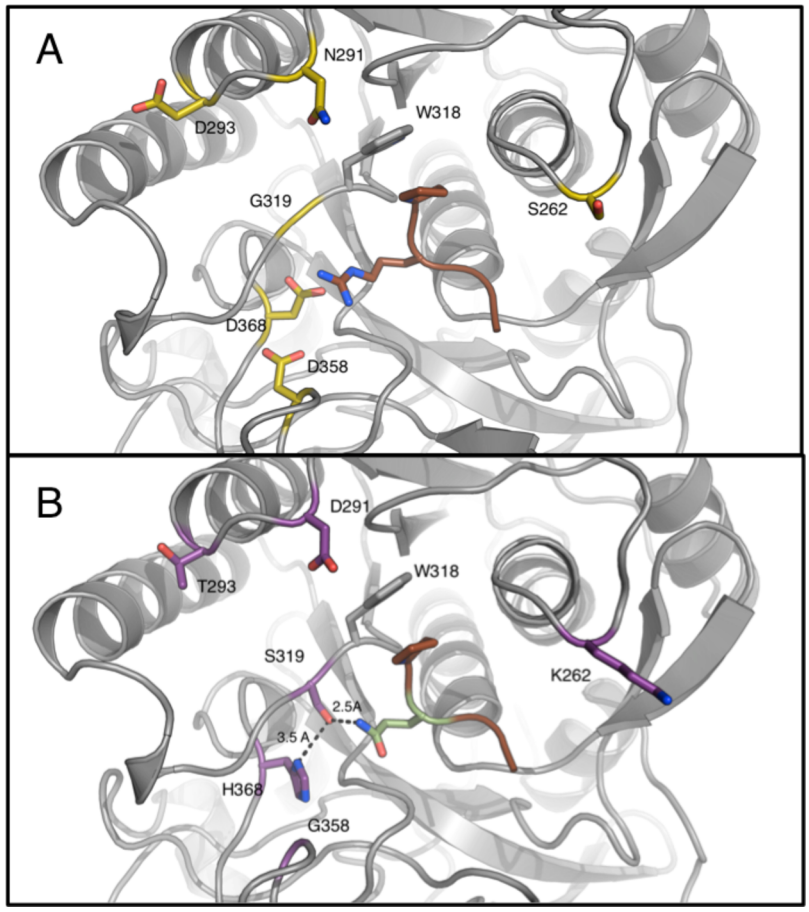


Figure 2. Computational models of the peptide binding sites for KumaWT and KumaMax. (A) KumaWT in complex with a PR dipeptide motif (brown). (B) KumaMax in complex with the designed PQ dipeptide motif (brown, native P; green, designed Q). Computationally designed residues in the active site are labeled and highlighted in sticks (KumaWT, gold; KumaMax, purple). The modeled peptides were based on a bound structure of kumamolisin-AS (PDB ID 1T1E) and final structures were generated using the Rosetta Molecular Modeling Suite. Images were generated using PyMol v1.5 (http://www.pymol.org/).

bound in the active site, so the residues of this polypeptide were mutated using Rosetta to the PQLP tetrapeptide motif. A total of 75 residues within an 8 Å sphere of the tetrapeptide were targeted for mutagenesis during the design process. Combinatorial sets of mutations were analyzed for their predicted effect on the overall energy of the new enzyme−PQLP substrate complex. A mutation set was considered for experimental characterization if the predicted energy of the enzyme−substrate complex was not significantly higher than wild-type. To accommodate the smaller, neutral amino acid glutamine, we focused our design efforts on removing the negative charge of the S1 pocket during the design process, filling in open space that resulted from the replacement of the larger arginine with glutamine and providing hydrogen bonds to the amide functional group of the glutamine. This computational modeling yielded 261 designs containing from one to seven simultaneous mutations.

In order to test the activity of these designed peptidases against the PQLP motif, the desired mutations were incorporated into the KumaWT nucleotide sequence using site-directed mutagenesis, and mutant enzyme variants were produced in *E. coli* BL21(DE3) cells.[16] Enzyme variants were then screened for enzymatic activity in clarified whole cell lysates at pH 4 using the synthesized peptide analogue QXL520-PQPQLP-K(5-FAM)-NH2 (FQ). This substrate is an α-gliadin hexapeptide analogue conjugated to 5-carboxyfluorescein (5-FAM) at the C-terminus and to the nonfluorescent quencher QXL520 at the N-terminus. Thus, peptidase activity can be measured by the increased fluorescence resulting from the release of 5-FAM from the quencher. Of the 261 enzyme variants tested in this assay, 20% had decreased enzymatic function compared with KumaWT, 30% were similar in activity to KumaWT, and 50% had an increase in activity against this substrate (Supplementary Table 1 and Supplementary Figure 1, Supporting Information). Twenty-eight of the most promising enzyme variants that exhibited a 2−70- fold increase in activity in cell lysates were then purified in order to obtain an accurate comparison of enzymatic activity to that of KumaWT. After purification and correction for protein concentration, the catalytic efficiencies of these enzymes, as determined by their k_{cat}/K_M values, ranged from 2-fold to 120-fold more active than KumaWT (Supplementary Table 2, Supporting Information). The most active variant, which we termed KumaMax, was selected for further characterization.

KumaMax contains seven mutations from the wild-type amino acid sequence: V119D, S262K, N291D, D293T, G319S, D358G, and D368H (Figure 2B). Of these, G319S, D358G, and D368H introduce a new hydrogen bond with the desired glutamine residue at position P1. As modeled, the G319S mutation introduces a hydroxyl group that is located 2.5 Å from the carbonyl oxygen of the glutamine amide, potentially

dx.doi.org/10.1021/ja3094795 | J. Am. Chem. Soc. 2012, 134, 20513−20520

Table 1. Catalytic Efficiency (M^{-1} s^{-1}) of Peptide Substrates for KumaMax and KumaWT[a]

	Qu-PQPQLP-Fl	Suc-APQ-pNA	Suc-APR-pNA	Suc-APE-pNA	Suc-AQP-pNA
KumaWT	4.9 ± 0.2	NDA[b]	131.8 ± 3.8	4.0 ± 0.1	NDA[b]
KumaMax	568.5 ± 14.6	6.7 ± 0.4	NDA[b]	1.4 ± 0.2	NDA[b]

[a]The catalytic efficiency (k_{cat}/K_M, $M^{-1}s^{-1}$) for both KumaWT and KumaMax for the fluorescently (Fl) quenched (Qu) PQPQLP peptide observed from a linear fit of velocity vs substrate profiles as no saturation was observed up to 100 mM substrate. The fluorescence signal was quantified as described in Methods with a standard curve that accounted for substrate quenching of product fluorescence. The catalytic efficiency for the pNA-linked peptides was determined in a similar manner, and is described in the Methods. All fits had at least five independently measured rates with an R^2 greater than 0.9 and are shown in Supplementary Figures 2 and 3, Supporting Information. [b]No detectable activity.

contributing a new hydrogen bond that interacts with glutamine in the P1 pocket. The D368H mutation is predicted to stabilize the serine hydroxyl, and its position in the active site is in turn sterically allowed by the D358G mutation. In addition to introducing interactions with the glutamine, these mutations also remove the two acidic residues predicted to stabilize the positively charged arginine residue in the native KumaWT substrate (Figure 2A). The mutation V119D, which was unexpectedly incorporated during site directed mutagenesis, is located in the propeptide domain and therefore does not affect catalytic activity of the mature enzyme. The other three mutations do not make direct contacts with residues in the P2−P2′ pockets and therefore likely introduce interactions with other components of the hexapeptide, the fluorophore, or the quencher. It is clear that these mutations are important for the overall catalytic enhancement observed, because the G319S/D358G/D368H triple mutant alone demonstrated only a 7-fold increase in catalytic activity over KumaWT; roughly 17-fold lower than that determined for KumaMax (Supplementary Table 2, Supporting Information).

Kinetic Characterization and Substrate Specificity. The catalytic efficiencies for KumaMax and KumaWT against the FQ immunogenic gluten substrate analogue, as calculated from a velocity versus substrate profile over 6−100 μM substrate, were found to be 568 and 4.9 M^{-1} s^{-1}, respectively (Table 1 and Supplementary Figure 2, Supporting Information). These values are consistent with the observation that KumaMax demonstrates 116-fold higher enzymatic activity toward the FQ substrate in the initial activity screen mentioned above. No significant saturation of velocity at these substrate concentrations was observed, and therefore the individual kinetic constants k_{cat} and K_M could not be determined. This is not surprising because previous analyses of the kinetic constants of KumaWT report a K_M of 40 μM[14] on a related substrate. It is therefore not unexpected that no significant saturation is observed at substrate concentrations less than 100 μM.

To confirm that the specificity of KumaMax had indeed been altered to prefer the PQ dipeptide over the native PR dipeptide of KumaWT, four peptides in the general form succinyl-alanine-P2-P1-P1′ were provided as substrates for both enzymes in order to assess P2 and P1 specificity. These peptides contained the reporter p-nitroaniline (pNA) at the P1′ position, which specifically enables a spectrophotometric readout of peptide cleavage between the P1 and P1′ positions. The P1 and P2 positions of the four peptides tested were proline−glutamine (PQ), proline−arginine (PR), glutamine−proline (QP), and proline−glutamate (PE). Catalytic efficiencies were experimentally determined, and kinetic values were calculated for each substrate as reported in Table 1. As in the determination of catalytic activities against the FQ substrate, no saturation of activity on these peptides by KumaWT or KumaMax was observed over a range of 15 μM to 1 mM. This suggests that the pNA group may partially disrupt binding in the P1′ pocket, since a substrate concentration of 1 mM is well beyond saturation levels previously reported for alternative KumaWT substrates.[12]

In this specificity assay, KumaMax demonstrated the highest activity on the PQ substrate, the dipeptide motif it was designed to recognize. KumaMax exhibited no significant catalytic activity against the QP or PR substrates, and roughly 5-fold lower activity on the isosteric PE substrate relative to the PQ substrate (Table 1). Consistent with previous reports,[13] KumaWT exhibited its highest level of activity on the PR motif, with significantly lower levels of activity on the three other peptide substrates. While catalytic activity of KumaWT on the PQ dipeptide motif has previously been reported,[13] no significant activity on the PQ dipeptide substrate was observed in our experiments. This may be due to disruptive effects of pNA on the binding of this peptide to the enzyme active site.

As discussed previously, there are several enzymes currently being explored as OET candidates for celiac disease. Two of these enzymes are engineered forms of the prolyl endopeptidase SC PEP and the glutamine-specific endoprotease EP-B2.[17] To compare the catalytic efficiencies of these peptidases to that of KumaMax, the native SC PEP and EP-B2 enzymes were expressed in *E. coli* BL21(DE3) cells and purified, and their catalytic activities were assessed. SC PEP demonstrated a catalytic efficiency of 1.6 M^{-1} s^{-1} on the FQ gluten substrate analogue at pH 4, a roughly 350-fold lower level of activity on this substrate than KumaMax. At pH 4, SC PEP did not exhibit significant activity on any of the four pNA-linked peptide substrates, including QP. Although previous studies using similar pNA-linked peptides have demonstrated activity of SC PEP on these substrates, those assays were performed at a pH of 4.5 or higher.[17] We found that SC PEP demonstrated significant levels of activity on the QP substrate at pH 7, with a catalytic efficiency of 7645 M^{-1} s^{-1}, thereby confirming that this recombinant SC PEP was functional (Supplementary Figure 4, Supporting Information). This is consistent with previous literature reporting that SC PEP has low to negligible levels of catalytic activity in the pH range of the stomach and is thus only expected to be effective once α-gliadin peptides have reached the small intestine.[17,18]

For EP-B2, only very low levels of activity were detected on the FQ substrate at pH 4, and no activity on any of the four pNA peptide substrates was observed. This is inconsistent with previous reports of EP-B2 activity using comparable substrates.[19] EP-B2 is a difficult enzyme to purify, because it forms inclusion bodies in *E. coli* and requires refolding to obtain active enzyme. We were unable to obtain soluble protein using previously reported methods for the refolding of EP-B2,[19−21] and thus we ultimately employed an on-column refolding process that resulted in the production of soluble protein.

Although this soluble EP-B2 demonstrated the expected self-processing activity at pH 4[19] (Supplementary Figure 5, Supporting Information), the lack of activity of this enzyme suggests that it may not have refolded properly using our methods. This could be due to alternative N- and C-terminal purification tags arising from the use of different protein expression vectors and warrants further investigation.

Protease Stability. In addition to demonstrating catalytic activity at low pH, any protein therapeutic intended for use in the human digestive tract must be resistant to degradation by digestive proteases. Two of the most abundant proteases in the stomach and small intestine are pepsin and trypsin, respectively.[19] Pepsin has optimal proteolytic activity at the low pH range of the stomach, while trypsin is primarily active at the more neutral pH of the small intestine. To assess the resistance of KumaMax to degradation by these proteases, 0.2 mg mL^{-1} of KumaMax was incubated with each protease, in their respective optimal pH ranges, at a concentration of 0.1 mg mL^{-1}, which is a physiologically relevant concentration of both pepsin and trypsin.[18,22] SC PEP and EP-B2 were included as controls, because EP-B2 has been established to be resistant to pepsin but susceptible to trypsin, and SC PEP is susceptible to both proteases.[17,19] Each protein was incubated in the presence or absence of the respective protease for 30 min, after which the proteins were heat inactivated and the remaining non-proteolyzed fraction was determined using an SDS−PAGE gel (Figure 3, Supplementary Figure 6, Supporting Information).

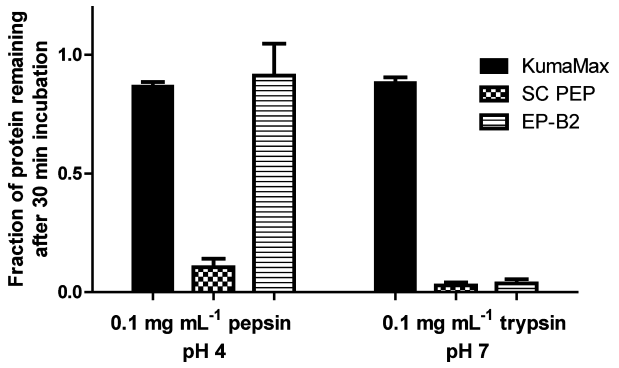


Figure 3. Protein stability after incubation with pepsin or trypsin. The fraction of intact protein after 30 min of incubation in the presence or absence of pepsin (at pH 4) or trypsin (at pH 7) was determined by quantification on a SDS−PAGE gel. Each protein was measured in triplicate, and the error bars represent the standard deviation. Protein gels are shown in Supplementary Figure 6, Supporting Information. Quantification was performed in ImageJ.

KumaMax had high stability upon incubation with either pepsin or trypsin, with roughly 90% of the protein remaining intact after a 30-min incubation with either protease (Figure 3). Consistent with previous reports, SC PEP was susceptible to both pepsin and trypsin, with less than 20% of the enzyme remaining after incubation with these proteases. As expected, trypsin efficiently proteolyzed EP-B2 with less than 10% enzyme remaining after incubation, but no significant degradation of EP-B2 was observed in the presence of pepsin. To confirm that observed protein degradation was due to protease activity and not to enzymatic self-processing, each enzyme was incubated at both pH 4 and 7, and apparent self-proteolysis was analyzed in the absence of other proteases over the course of an hour (Supplementary Figure 3, Supporting Information). As expected, KumaMax and EP-B2, but not SC PEP, demonstrated self-processing of the pro-enzyme to form the active enzyme in fewer than 10 min at pH 4, after which all three proteins remained >90% stable over the course of the hour. None of the proteins showed significant levels of self-processing or proteolysis during incubation for 1 h at pH 7 (Supplementary Figure 5, Supporting Information).

Degradation of an Immunogenic α9-Gliadin Peptide. The significant level of catalytic activity and substrate specificity on immunogenic peptide analogues (Table 1) is promising for the use of KumaMax as a therapeutic in OET for gluten intolerance. However, these assays do not directly assess the ability of KumaMax to degrade relevant immunogenic peptides derived from gluten. Therefore, we examined the direct proteolytic activity of this enzyme toward an immunodominant peptide present in α9-gliadin, QLQPFPQPQLPY.[23,24]

KumaMax was incubated at 37 °C in pH 4 sodium acetate with 500 μM of the α9-gliadin peptide at a roughly 1:100 enzyme to peptide molar ratio, which is a physiologically relevant concentration of this peptide in the human stomach (Supporting Information). SC PEP was included in this experiment for the sake of comparison, since this enzyme has significantly less activity against the FQ substrate than KumaMax at pH 4. Samples from the incubation were quenched every 10 min in 80% acetonitrile to halt the proteolysis reaction. The remaining fraction of intact immunogenic peptide was determined using ultrahigh-performance liquid chromatography−mass spectrometry, in which the M + H parent ion of the α9-gliadin peptide was monitored. KumaMax demonstrated a high level of activity against the immunogenic peptide in this assay, as over 95% of the immunogenic peptide had been proteolyzed after a 50-min incubation with KumaMax, while no significant degradation of the peptide was observed in the presence of SC PEP or in the absence of a peptidase (Figure 4A). The half-life of the peptide in the presence of KumaMax was determined by plotting the fraction of peptide remaining against the incubation time and was calculated to be 8.5 ± 0.7 min (Figure 4B).

DISCUSSION

Computational enzyme design allows for the introduction of new chemical and catalytic traits into pre-existing enzyme scaffolds and thus holds great promise as a method to speed progress in the development of next-generation enzyme therapeutics. Celiac disease is a human condition that is an ideal model for the use of rational design to produce an enzyme therapeutic, due to the stringent set of requirements the disease sets forth for its potential treatments. Oral enzyme therapy is also an attractive method for the treatment of celiac disease since it does not require invasive methods of administration or rigorous monitoring of food intake. An ideal peptidase for use in OET would maintain activity in a pH range from 2 to 4 at 37 °C, resist degradation by common digestive proteases, and be specific for the common proline−glutamine motif found in immunogenic gluten-derived peptides. In addition, the protein should be easily produced using recombinant methods. While it is unlikely that a single natural enzyme will have all of these properties, we demonstrate that a naturally occurring protein with several of these important characteristics can be engineered to build in the others.

The engineered endopeptidase, KumaMax, has a high level of activity and specificity toward the desired PQ dipeptide motif

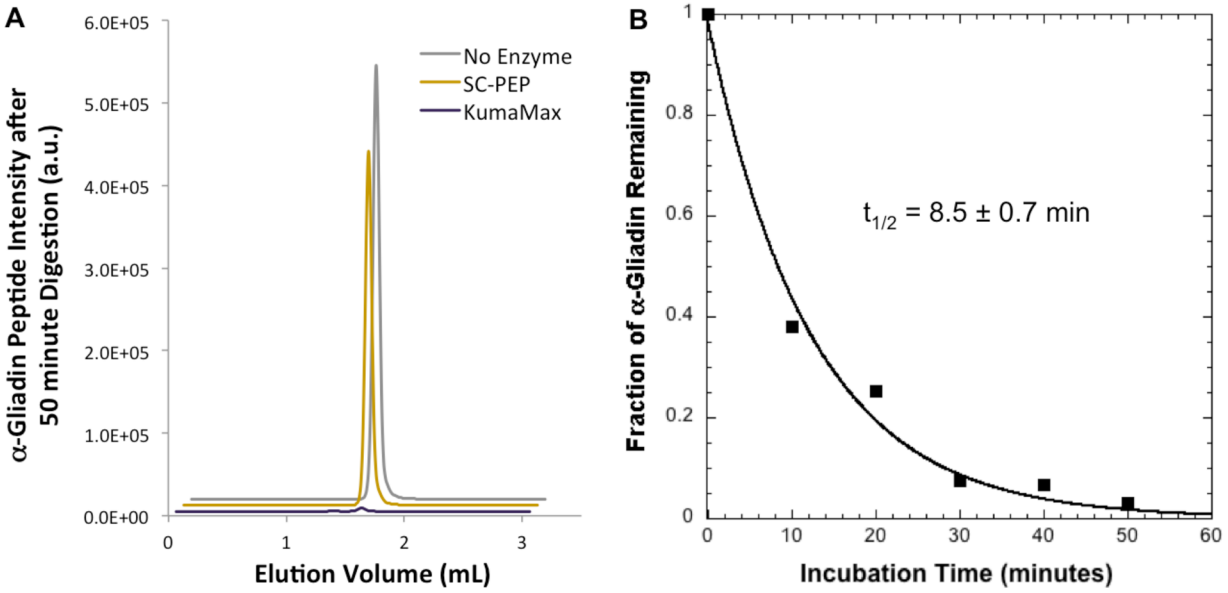


Figure 4. Immunogenic α9-gliadin peptide degradation by KumaMax. (A) Reaction chromatograms measuring the abundance of the M + H ion of the parent α9-gliadin peptide after 50 min of incubation with no enzyme (gray), SC PEP (gold), or KumaMax (purple). (B) The fraction of α9-gliadin peptide remaining in the presence of KumaMax as a function of incubation time at pH 4. The curve is a sample exponential fitting. The R^2 value was 0.97.

(Table 1). The specificity for the PQ motif, compared with the specificity of the native KumaWT for the PR motif, potentially derives from the addition of new hydrogen bonds in the S1 pocket of KumaMax that, as modeled, make direct molecular contacts with the glutamine in this dipeptide motif (Figure 2B). This specificity switch not only directs activity against a motif found commonly throughout gluten but also greatly decreases activity against nontargeted substrates (Table 1). The ability of an orally administered peptidase to specifically recognize its target substrates is an important characteristic because it reduces competitive inhibition by the large number of other peptides produced in the stomach during digestion of a meal and serves to decrease off-target digestion of other necessary gut proteins. While most immunogenic peptides derived from gluten contain the PQ motif, it may be beneficial in OET for gluten intolerance to include a set of enzymes with altered substrate specificities that specifically target other gluten-derived motifs. KumaMax or KumaWT can potentially act as platforms for engineering greater specificity, because KumaWT has demonstrated some level of selectivity beyond the P2 and P1 sites.[13] With the methods described here, a panel of customized peptidases specific for unique immunogenic epitopes could be generated, and this is currently being explored.

KumaMax catalytic activity against the PQ motif can likely be further optimized, since we tested only a small fraction of the possible combinatorial active site mutations. The requirement for activity at low pH precludes the ability to screen for enzyme activities using commonly implemented high throughput assays, such as genetic complementation.[25−27] The development of new computational tools for the design of directed libraries would therefore enable further optimization of catalytic activity. In addition, characterizing the effects of the mutations made to KumaMax on the pH profile and thermal stability of this enzyme may provide insight into the direction of further optimization strategies.

CONCLUSION

In this study, we describe the use of computational tools to design a novel enzyme that is specific for the proteolysis of a dipeptide motif found in immunogenic peptides that elicit gluten-derived pathology in patients with celiac disease. This computational method enabled the identification of a set of mutations that resulted in a 116-fold increase in enzyme activity on the desired substrate and an 877-fold switch in substrate specificity. We demonstrated that this enzyme has great promise as a potential gluten-degrading therapeutic due to its resistance to proteolysis and its ability to degrade over 95% of an immunogenic peptide under physiologically relevant conditions. Future work focused on the further characterization of this designed enzyme will ultimately shed light on its suitability for use as a therapeutic. The continued enhancement of catalytic activity using this platform and its expansion to novel specificities could lead to a new generation of protein therapeutics for celiac disease and related food intolerances.

METHODS

General Materials. Sequences coding for KumaWT and EP-B2 were synthesized by Genscript to be codon-optimized for *E. coli* and cloned into a pET29b+ vector. SC PEP was a generous gift from Chaitan Khosla. All peptide substrates were purchased from Anaspec, and biochemical reagents and materials were purchased from Sigma or Fisher.

Enzyme Mutagenesis, Expression, Screening, and Purification. Site-directed mutagenesis was performed according to the protocol developed by Kunkel to generate mutations to KumaWT,[28] and variants were expressed and harvested in a clarified whole-cell lysate (see Supporting Information for details). Clarified lysate harboring mutant enzymes was screened for activity against the FQ substrate by adding 10 μL of cell lysate supernatant to 90 μL of 5 μM substrate in a 96-well black plate, and the fluorescence was measured at 30-s intervals for 60 min with a 455 nm wavelength for excitation and a 515 nm wavelength for emission.

Variants of KumaWT that displayed a 2−70-fold increase in activity were expressed and purified using 1 mL TALON cobalt affinity columns (see Supporting Information for details). Protein concen-

dx.doi.org/10.1021/ja3094795 | J. Am. Chem. Soc. 2012, 134, 20513−20520

trations were calculated spectrophotometrically with extinction coefficients calculated from their respective primary sequences of 53985 M^{-1} cm^{-1} for KumaWT and all KumaWT variants, 152290 M^{-1} cm^{-1} for SC PEP, and 58245 M^{-1} cm^{-1} for EP-B2.

Kinetic Characterization. The activity of computationally designed enzyme variants against the PQ motif was measured by hydrolysis of the fluorescently quenched α-gliadin hexapeptide analogue QXL520-PQPQLP-K(5-FAM)-NH2 (FQ) as a substrate (5-FAM, 5-carboxyfluorescein; QXL520, a proprietary fluorescence quencher produced by Anaspec). Each enzyme was incubated at room temperature for 15 min in 100 mM pH 4 sodium acetate buffer. After 15 min, 50 mL of fluorescent substrate (50, 25, 12.5, 6.3, or 0 μM) was incubated with KumaMax (0.05 μM), KumaWT (0.5 μM), SC PEP (0.5 μM), and EP-B2 (0.5 μM). Fluorescence intensities were used as a readout of the activities of these enzymes and were determined using a SpectraMax M5e spectrophotometer (Molecular Devices) over 60 min, using 455 nm wavelength for excitation and reading 515 nm wavelength for emission.

Purified enzymes were also tested for their specificity toward different dipeptide motifs using a variety of chromogenic substrates that release pNA upon hydrolysis: [Suc-APQ-pNA], [Suc-AQP-pNA], [Suc-APE-pNA], and [Suc-APR-pNA]. Each enzyme was incubated at room temperature for 15 min in 100 mM pH 4 sodium acetate buffer. After 15 min, 20 μL of substrate were added to the incubated enzyme such that the final substrate concentrations were 1000, 500, 250, 125, 62.5, 31.3, 15.6, or 0 μM. All enzymes tested in this assay were at a final concentration of 0.5 μM. Released pNA was quantified by monitoring absorption at 385 nm using a SpectraMax M5e spectrophotometer (Molecular Devices) over the course of an hour.

To correct for intermolecular fluorescence quenching in samples that contained high levels of substrate, a standard product curve was generated for each substrate concentration. Substrate concentrations were 100, 50, 25, 12.5, 6.3, and 0 μM, and product concentrations were 20, 5, 1.3, 0.3, 0.08, and 0 μM (Supplementary Figure 7, Supporting Information). The standard curve for the absorbent *p*-nitroaniline-linked peptides contained product concentrations of 100, 50, 25, 12.5, 6.3, 3.1, 1.6, 0.8, 0.4, 0.2, 0.1, and 0 μM diluted in pH 4 buffer (Supplementary Figure 8, Supporting Information).

Protease Stability. Enzyme stability was determined in the presence of the digestive proteases pepsin and trypsin. KumaMax, SC PEP, and EP-B2 were incubated in 100 mM sodium acetate pH 4 buffer (for pepsin digestion assays) or pH 7 dialysis buffer (20% glycerol, 50 mM HEPES pH 7, 500 mM NaCl, 1 mM β-mercaptoethanol) used during purification of the protein (for trypsin digestion assays). Each enzyme was preincubated at 37 °C for 15 min in the appropriate buffer at a concentration of 0.2 mg mL^{-1}. After preincubation, 0.1 mg mL^{-1} of trypsin or pepsin was added. The reactions were incubated at 37 °C for 30 min, after which the reactions were quenched by the addition of 2% SDS and boiling for 5 min. These samples were then analyzed on an SDS–PAGE gel, and protein bands were quantified using ImageJ.

The rate of protein autoprocessing was determined at pH 4 and 7 in the absence of pepsin or trypsin. Each enzyme, at a concentration of 0.2 mg mL^{-1}, was incubated in pH 4 100 mM sodium acetate or pH 7 dialysis buffer. Aliquots were taken at 20, 40, and 60 min, quenched with 2% SDS, and boiled for 5 min. Samples were run on an SDS–PAGE gel, and protein concentrations were quantified using ImageJ.

Computational Modeling. The structure of the intact propeptide form of KumaWT (PDB-ID 1T1E) was used to identify where the P2–P2′ peptide bindings sites are located. The tetrapeptide spanning the P2–P2′ site in 1T1E was copied into a structure of a processed apoprotein of KumaWT (PDB-ID 1SIO). Catalytic constraints were used to define a 90° angle of attack of the serine 278 hydroxyl on carbonyl of the P1–P1′ amide bond, at a distance of 2.0 Å between the oxygen and carbon and 109° off of the hydroxyl.[29] The structure was then loaded into Foldit, a graphical user interface to the Rosetta Molecular Modeling Suite. While keeping the main chain backbone fixed, the side chain of the tetrapeptide was converted into the sequence proline (P2), glutamine (P1), leucine (P1′), and proline (P2′). This interface was repacked and minimized in Foldit, optimizing both amino acid side chain and backbone conformations in the presence of the catalytic constraints. With the minimized structure, mutations surrounding the predicted P2–P2′ peptide binding site were modeled and evaluated in Foldit to identify favorable subsets of mutations that made novel interactions with the P1 glutamine residue. As discussed in the Results section, mutation sets were considered for experimental characterization if the predicted energy was either reduced relative to the native substrate or was significantly higher than wild-type. In addition, mutations from small amino acids to large amino acids were prioritized in order to fill open space that resulted from the replacement of the larger arginine with glutamine in the P1 position.

LCMS Gliadin Degradation Assay. Enzyme activity on full-length α9-gliadin was measured using ultrahigh-performance liquid chromatography–mass spectrometry (UHPLC–MS). For each enzyme, 7 μL of 1 M sodium acetate buffer, pH 4, was added to 28 μL of 5 μM enzyme (or 28 μL dialysis buffer as a negative control), and incubated alongside separate tubes of 3 μL of α9-gliadin, also at pH 4, at 37 °C for 15 min. After incubation, 27 μL of each enzyme mixture was added to the preincubated α9-gliadin. The enzyme–gliadin mixtures were incubated at 37 °C, and 5 μL samples were taken at 10, 20, 30, 40, and 50 min. Each time point sample was quenched in 95 μL of 80% acetonitrile with 1% formic acid and 33 μM leupeptin as an internal control. This solution was incubated for 5 min, and the precipitated protein was then removed by filtration using a Millipore multiscreen solvinert filter plate (product number MSRLN0450). Five microliters of the filtered quench solution was subsequently injected and analyzed using an UHPLC–MS assay. The column used for the chromatography run was a Hypersil Gold C18 (dimensions 100 mm × 2.1 mm, 1.9 μm particle size, Thermo). The following gradient was performed for the chromatography run at a flow of 500 μL min^{-1}: 95:5 water/ACN for 30 s, followed by a gradient over 4.5 min ending at 5:95 water/ACN. The column was then restored to the initial 95:5 water/ACN buffer and allowed to equilibrate in this buffer for 1 min before the next injection. All buffers contained 0.1% formic acid. The M + H ion for α9-gliadin (1456.7) peptide and leupeptin (427.5, used as an internal standard for injection differences and sample evaporation) were detected using a TSQ Quantum Access (Thermo Fisher Scientific) with an ESI probe.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

List of all of the computationally designed mutations assayed, data relating to kinetics of the engineered enzymes, pH and protease stability, and standard product curves for KumaWT, KumaMax, SC PEP, and EP-B2, detailed methods for the LCMS assay of gluten degradation, enzyme mutagenesis and screening, and protein expression and purification, and protein and nucleotide sequences for KumaWT, KumaMax, and EP-B2. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

jbsiegel@ucdavis.edu; iswanson@u.washington.edu

Author Contributions

‡These authors contributed equally.

Notes

The authors declare the following competing financial interest(s): Ingrid Pultz, Justin Siegel and David Baker are founders of Proteus Biologics, a company formed around the technology presented in this publication.

■ ACKNOWLEDGMENTS

We thank the International Genetically Engineered Machine (iGEM) Competition for promoting undergraduate research

projects and providing a rich academic environment. Alan Weiner, Dominic Chung, and Ling Lin Liu (Department of Biochemistry, University of Washington) generously provided laboratory space for the undergraduate students involved in this work. We would also like to thank Chaitan Khosla for his insightful comments on the direction for this research, as well as for providing the genetic construct encoding SC PEP. This work was supported by the Howard Hughes Medical Institute and a grant from the Defense Advanced Research Projects Agency.

REFERENCES

(1) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science* **2010**, *329*, 309.
(2) Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E. M.; Wilson, I. A.; Baker, D. *Science* **2011**, *332*, 816.
(3) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364.
(4) Sollid, L. M. *Nat. Rev. Immunol.* **2002**, *2*, 647.
(5) Wieser, H. *Food Microbiol.* **2007**, *24*, 115.
(6) Shan, L.; Molberg, O.; Parrot, I.; Hausch, F.; Filiz, F.; Gray, G. M.; Sollid, L. M.; Khosla, C. *Science* **2002**, *297*, 2275.
(7) Gardner, J. D.; Ciociola, A. A.; Robinson, M. *J. Appl. Physiol.* **2002**, *92*, 427.
(8) Shan, L.; Qiao, S.-W.; Arentz-Hansen, H.; Molberg, Ø.; Gray, G. M.; Sollid, L. M.; Khosla, C. *J. Proteome Res.* **2005**, *4*, 1732.
(9) Shan, L.; Marti, T.; Sollid, L. M.; Gray, G. M.; Khosla, C. *Biochem. J.* **2004**, *383*, 311.
(10) Siegel, M.; Bethune, M. T.; Gass, J.; Ehren, J.; Xia, J.; Johannsen, A.; Stuge, T. B.; Gray, G. M.; Lee, P. P.; Khosla, C. *Chem. Biol.* **2006**, *13*, 649.
(11) Stepniak, D.; Spaenij-Dekking, L.; Mitea, C.; Moester, M.; de Ru, A.; Baak-Pablo, R.; van Veelen, P.; Edens, L.; Koning, F. *Am. J. Physiol.: Gastrointest. Liver Physiol.* **2006**, *291*, G621.
(12) Okubo, A.; Li, M.; Ashida, M.; Oyama, H.; Gustchina, A.; Oda, K.; Dunn, B. M.; Wlodawer, A.; Nakayama, T. *FEBS J.* **2006**, *273*, 2563.
(13) Wlodawer, A.; Li, M.; Gustchina, A.; Tsuruoka, N.; Ashida, M.; Minakata, H.; Oyama, H.; Oda, K.; Nishino, T.; Nakayama, T. *J. Biol. Chem.* **2004**, *279*, 21500.
(14) Oda, K.; Ogasawara, S.; Oyama, H.; Dunn, B. M. *J. Biochem.* **2000**, *128*, 499.
(15) 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.; 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.; Bradley, P. *Methods Enzymol.* **2011**, *487*, 545.
(16) Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Cooper, S.; Khatib, F.; Shen, B. W.; Players, F.; Stoddard, B. L.; Popovic, Z.; Baker, D. *Nat. Biotechnol.* **2012**, *30*, 190.
(17) Ehren, J.; Govindarajan, S.; Moron, B.; Minshull, J.; Khosla, C. *Protein Eng., Des. Sel.* **2008**, *21*, 699.
(18) Bethune, M. T.; Khosla, C. *Methods Enzymol.* **2012**, *502*, 241.
(19) Bethune, M. T.; Strop, P.; Tang, Y.; Sollid, L. M.; Khosla, C. *Chem. Biol.* **2006**, *13*, 637.
(20) Vora, H.; McIntire, J.; Kumar, P.; Deshpande, M.; Khosla, C. *Biotechnol. Bioeng.* **2007**, *98*, 177.
(21) Gass, J.; Vora, H.; Bethune, M. T.; Gray, G. M.; Khosla, C. *J. Pharmacol. Exp. Ther.* **2006**, *318*, 1178.
(22) Gass, J.; Bethune, M. T.; Siegel, M.; Spencer, A.; Khosla, C. *Gastroenterology* **2007**, *133*, 472.
(23) Arentz-Hansen, H.; Korner, R.; Molberg, O.; Quarsten, H.; Vader, W.; Kooy, Y. M.; Lundin, K. E.; Koning, F.; Roepstorff, P.; Sollid, L. M.; McAdam, S. N. *J. Exp. Med.* **2000**, *191*, 603.
(24) Hausch, F.; Shan, L.; Santiago, N. A.; Gray, G. M.; Khosla, C. *Am. J. Physiol.: Gastrointest. Liver Physiol.* **2002**, *283*, G996.
(25) Arnold, F. H. *Nature* **2001**, *409*, 253.
(26) Romero, P. A.; Arnold, F. H. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866.
(27) Bershtein, S.; Tawfik, D. S. *Curr. Opin. Chem. Biol.* **2008**, *12*, 151.
(28) Kunkel, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 488.
(29) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. *PLoS One* **2011**, *6*, No. e19230.