

Computer-aided design of a catalyst for Edman degradation utilizing substrate-assisted catalysis

Benjamin Borgo¹ and James J. Havranek^{2*}

¹Program in Computational and Systems Biology, Washington University in St. Louis, St. Louis, Missouri 63110

²Department of Genetics, Washington University in St. Louis, St. Louis, Missouri, 63110

Received 1 October 2014; Accepted 23 December 2014

DOI: 10.1002/pro.2633

Published online 16 December 2014 proteinscience.org

Abstract: Molecular biology has been revolutionized by the miniaturization and parallelization of DNA sequencing assays previously performed on bulk samples. Many of these technologies rely on biomolecular reagents to facilitate detection, synthesis, or labeling of samples. To aid in the construction of analogous experimental approaches for proteins and peptides, we have used computer-aided design to engineer an enzyme capable of catalyzing the cleavage step of the Edman degradation. We exploit the similarity between the sulfur nucleophile on the Edman reagent and the catalytic cysteine in a naturally occurring protease to adopt a substrate-assisted mechanism for achieving controlled, step-wise removal of N-terminal amino acids. The ability to expose amino acids iteratively at the N-terminus of peptides is a central requirement for protein sequencing techniques that utilize processive degradation of the peptide chain. While this can be easily accomplished using the chemical Edman degradation, achieving this activity enzymatically in aqueous solution removes the requirement for harsh acid catalysis, improving compatibility with low adsorption detection surfaces, such as those used in single molecule assays.

Keywords: protein engineering; substrate-assisted catalysis; Edman degradation; peptide analysis

Introduction

Technologies for the high-throughput, massively parallel analysis of proteins and peptides lag behind those of next-generation DNA sequencing. DNA analysis has benefited from the chemical and structural uniformity of nucleic acids, which simplifies analyte immobilization and detection. Arguably, the most important advantage that DNA assays hold

over peptide assays is the ready availability of naturally occurring enzymes that, with little or no modification, are sufficient to enable sequencing by synthesis, by ligation, and via pyrosequencing.¹ A significant challenge for any peptide sequencing assay seeking to harness the miniaturization and parallelization advantages of next-generation technologies is to engineer molecular tools for the detection and processing of peptide samples.

The classic Edman degradation reaction provides a controlled method for removing individual amino acids from the N-terminus of a peptide chain,^{2,3} and would benefit from the development of an enzymatic catalyst. The Edman degradation consists of two steps. In the first, the Edman reagent, phenylisothiocyanate (PITC), is coupled to the N-terminus of a peptide, forming a phenylthiocarbonyl (PTC) derivative. This step proceeds in mildly basic conditions, and is unlikely to prove problematic in future peptide assays. In the second step, the

This is an open access article under the terms of the Creative Commons Attribution NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Grant sponsor: National Institute of General Medical Sciences of the National Institutes of Health; Grant number: R01GM101602.

*Correspondence to: James J. Havranek, Campus Box 8232, 4566 Scott Avenue, St. Louis, MO 63110.

E-mail: havranek@genetics.wustl.edu

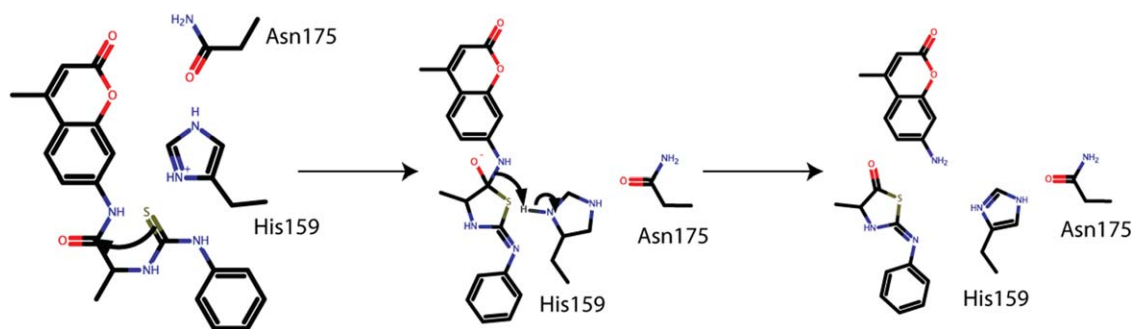


Figure 1. Intended mechanism for substrate assisted catalysis using a mutated cysteine protease. The proposed mechanism for Edmanase is modeled on the mechanism of cysteine proteases. The substrate shown is PTC-Ala-AMC, which is used for all of the control experiments. In the left side, His159 serves to stabilize the sulfur atom in the PTC group as it performs nucleophilic attack upon the carbon in the Ala-AMC amide bond. Next, His159 serves as a base, donating a proton to resolve the tetrahedral intermediate and liberating the AMC group. After cleavage is complete, His159 accepts a proton from the solvent, completing the catalytic cycle.

first amino acid of the peptide is cleaved, with the sulfur atom of the PTC group acting as a nucleophile. The power of this reaction lies in the regeneration of a native-like N-terminus, permitting multiple rounds of analysis. Unfortunately, the harsh conditions utilized in the cleavage step of the reaction (anhydrous trifluoroacetic acid) may irreversibly damage the surfaces used for immobilization and detection of macromolecules, especially when multiple rounds of cleavage are required. An enzymatic route to the cleavage step of Edman degradation is desirable because it could accomplish step-wise removal of amino acids in neutral, aqueous buffers that are benign to both the detection apparatus and other protein components required for the experimental workflow. However, this reaction is not a part of the biochemical repertoire of any organism, and as a result we must turn to protein engineering to provide this activity.

A promising strategy for engineering an enzyme for the cleavage step of the Edman reaction (which we will term an “Edmanase”) is substrate-assisted catalysis.⁴ In this approach, an enzyme is inactivated by mutation of a key catalytic residue, but activity is restored when the substrate can provide a similar functional group. For example, when the serine residue of a catalytic triad⁵ is mutated in a serine protease, then reintroduced in a peptide substrate, a highly specific site for proteolytic cleavage is generated.⁶ Substrate assisted catalysis has been used by several groups to engineer more specific variants of proteases.^{7,8} This approach is an excellent fit for the Edman cleavage reaction because the Edman reagent, phenylisothiocyanate (PITC), provides a sulfur nucleophile that becomes part of the substrate after the coupling step of Edman degradation. Peptidases utilizing catalytic cysteines are ideal candidate scaffolds for engineering an Edmanase enzyme via substrate-assisted catalysis, due to the chemical similarity of their nucleophiles and the straight-forward mutational route to removal of the catalytic sulfhydryl group.

In this manuscript, we describe the computer-aided design of an Edmanase. We selected a cysteine protease scaffold, and reengineered it using computer-aided design to catalyze the Edman degradation reaction via a substrate-assisted mechanism. The resulting enzyme exhibits improved cleavage activity for amino acid–amidomethylcoumarin (AMC) substrates after coupling of the Edman reagent at the N-terminus, and catalyzes the removal of the N-terminal derivatized amino acid with catalytic efficiencies (k_{cat}/K_M) of up to $10^4 \text{ s}^{-1} \text{ M}^{-1}$. Additionally, it exhibits activity for a range of amino acids at the P1 position of the substrate, suggesting that the enzyme can substitute for the standard acid catalysis in the cleavage step of the Edman reaction. We expect that the ability to carry out the Edman reaction under biocompatible conditions will prove useful for the development of next-generation peptide analysis assays.

Results

Selection of a template for engineering an Edmanase

We desired several properties for a design scaffold for engineering an Edmanase. First, our substrate-assisted catalysis strategy led us to consider only cysteine proteases. These enzymes are more likely to be compatible with the sulfur nucleophile on the substrate, and suggest a straightforward catalytic mechanism (Fig. 1). Second, we required that the protease have little to no preference for the P1 position of the substrate. This aids in eliminating dependence of catalytic activity on the side-chain identity at this position. Third, we sought a protease that could tolerate an aromatic residue at the P2 position of the substrate. Structural modeling suggested that the phenyl moiety of the Edman reagent had the potential to mimic a phenylalanine side chain at this position. Finally, we required that a crystal structure for any candidate scaffold be available in the protein databank (PDB),⁹ preferably in complex with a competitive inhibitor.

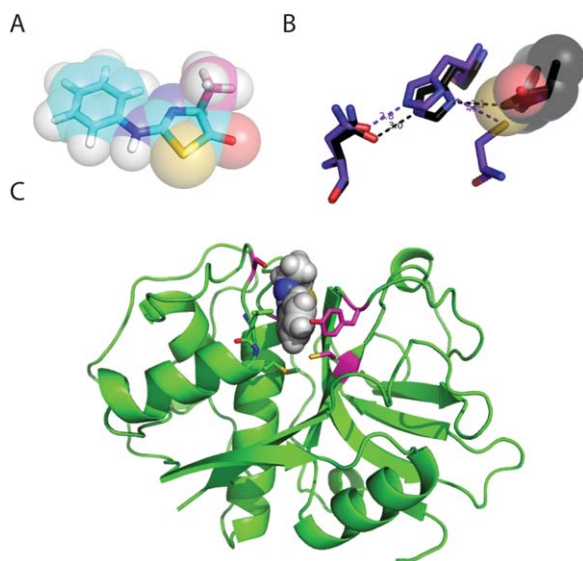


Figure 2. Model for the post-cleavage product of Edman degradation. (A) A model for the small molecule corresponding to a PITC-derivatized N-terminal alanine residue after cleavage, but before rearrangement into the phenylthiohydantoin derivative, was predicted using the quantum chemistry program GAMESS. The methyl group that corresponds to the side chain of alanine is shown in magenta. (B) The product model is shown in relation to the active site residues when modeled in complex with cruzain. The catalytic triad residues of cruzain are shown as purple sticks. In the intended catalytic mechanism for Edmanase, the sulfur atom in the PITC group takes the place of the sulfhydryl nucleophile, while the other two members of the triad (shown in black) are retained to activate the sulfur for nucleophilic attack on the N-terminal peptide bond. (C) The product analog is shown in its docked location. The mutations to cruzain are shown in magenta. The methyl group of the molecule corresponds to the side chain of an alanine residue, and is directed into solvent and away from the active site.

These criteria led us to select the cysteine protease cruzain from *Trypanosoma Cruzei* (PDB code: 1AIM¹⁰). As desired, the S1 pocket of the protease contributes little to the specificity of the enzyme.¹⁰ This protease is also known to prefer phenylalanine at the P2 position of the substrate.¹¹ Multiple crystal structures have been solved of this protein, including multiple co-crystals with a diverse set of inhibitors.

Computer-aided redesign of Edmanase binding pocket

We used the quantum chemical package GAMESS¹² to create a model of the immediate product of the cleavage step of the Edman degradation for an N-terminal alanine substrate [Fig. 2(A)]. As a starting template for protein design, we selected a high-resolution (2.0 Å) model of cruzain in complex with a covalent inhibitor (PDB code: 1AIM¹⁰). To prepare the cruzain scaffold for computer modeling, we removed the inhibitor and mutated the active site

nucleophile to glycine (C25G) *in silico* to create a cavity in the active site. Previous studies have shown that mutation of the active-site cysteine reduces activity to background levels (at least a million-fold).¹³ We then docked our model for the cleavage product into the binding pocket, constraining the sulfur in the model to occupy the same location previously occupied by the sulfhydryl group of Cys25 [Fig. 2(B)]. Using the Rosetta full atom energy function¹⁴ and side chain prediction algorithms to guide us, we manually selected three additional mutations (G65S, A133C, and L157Y) [Fig. 1(C)]. After constrained relaxation, these mutations improved the predicted binding energy for the product model by 13.8 Rosetta Energy Units (REUs), and resulted in a predicted improvement in steric complementarity between the substrate and the catalytic site (Fig. 3).

These three mutations [Fig. 4(A)] were inspired by commonly occurring, favorable interactions involving the chemical moieties found in our small molecule model, a strategy that previously allowed us to reengineer the specificity of methionine aminopeptidase.¹⁵ A mutation of leucine to tyrosine at position 157 is intended to introduce a favorable “ring-to-edge” interaction between the tyrosine side chain and the phenyl group on the substrate.^{16–18} Additionally, the model suggests that it can cap the binding pocket [Fig. 3(B)], which may help to maintain the position of the sulfur atom in the substrate near the original position of the catalytic cysteine, as seen in the alignment of the design with the native catalytic triad [Fig. 4(B)]. The introduction of an electron-rich cysteine at position 133 also introduces a canonical interaction with the lower edge of the aromatic ring.¹⁹ The final mutation (G65S) is predicted to satisfy a solvent-exposed hydrogen bond on the substrate. The resulting design model exhibits tight packing around the modeled substrate, and places the sulfur nucleophile in the correct orientation to be activated by the remaining components of the catalytic triad.

Edmanase exhibits specificity for the Edman reagent

We expressed and characterized the four-fold mutant using an assay that follows the accumulation of fluorescence as Edmanase cleaves the PTC-amino acid from the “N-terminus” of a PTC-Ala-AMC substrate (see Methods section). As negative controls, we also performed the assay in the presence of BSA instead of Edmanase and using substrate that had not been derivatized with PITC (denoted PITC- in Fig. 5). Low levels of fluorescence accumulation are observed with BSA (Fig. 5, BSA PITC- and BSA PITC+), and are attributed to the uncatalyzed rate of Ala-AMC bond cleavage under mild aqueous conditions. Cleavage of Ala-AMC (PITC-) substrate by Edmanase is comparable to that of the PITC+

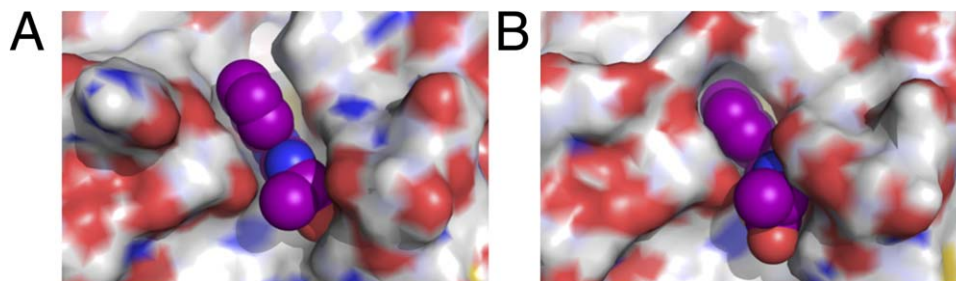


Figure 3. Redesign of the binding pocket of cruzain. (A) A predicted model for the post-cleavage PTC-Ala molecule (rendered in magenta) is shown docked into the binding pocket of the protein cruzain (PDB code: 1AIM¹⁰). In the surface representation of the protein, acidic and basic amino acids are rendered in red and blue, respectively. This protein was selected as the design template in part for its ability to accommodate the small molecule. (B) A predicted model for the redesigned enzyme in complex with the hypothetical PTC-Ala product shows enhanced steric complementarity. The improvements result from computer-aided redesign of four residues in the active site and surrounding pocket, along with constrained optimization of the small molecule docking position. The methyl group corresponding to the side chain of an alanine residue is directed away from the binding pocket.

substrate with BSA. There is a pronounced increase in catalytic efficiency for the PITC+ substrate compared to PITC- in the presence Edmanase of 282-fold (Fig. 5, Table I). Thus, robust cleavage of the Ala-AMC bond requires both derivatization of the

small molecule with PITC and the presence of Edmanase.

Edmanase exhibits enzymatic activity for a diverse set of P1 residues

We next investigated whether or not the identity of the amino acid at the P1 position of the substrate had an effect on catalytic efficiency. Our model orients the amino acid side chain of the substrate away from the body of the protein [Fig. 3(B)]. As expected based on the predicted structure of our design, cleavage occurred regardless of which amino-acid side-chain was present in the fluorescent substrate (Fig. 6). The catalytic efficiency did vary significantly, however, from $\sim 10^4$ s⁻¹ M⁻¹ for aspartate and

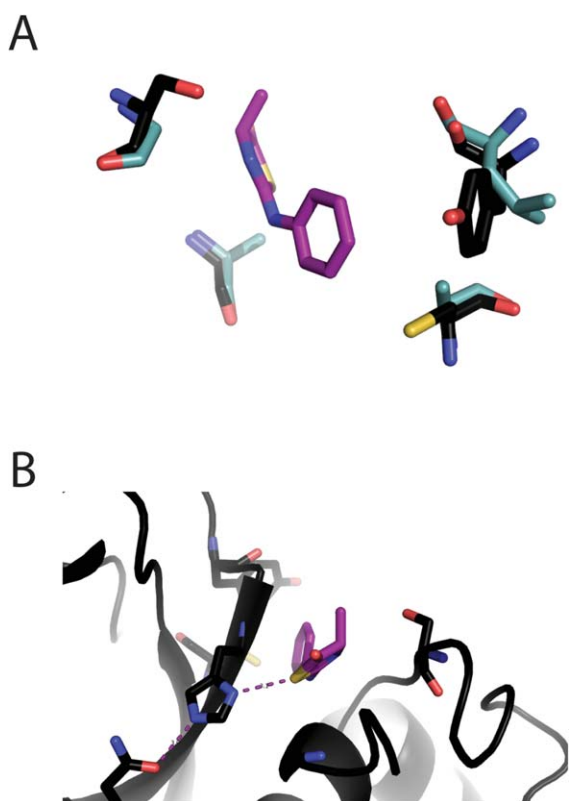


Figure 4. Four mutations alter the activity of cruzain to catalyze the Edman degradation. (A) The computationally-guided mutations (black) are shown over-laid on the original cruzain residues (cyan). (B) The modeled mutations are shown in the context of the cruzain scaffold showing the catalytic triad using the substrate-assisted mechanism. The modeled structure demonstrates that the placement of the substrate enforced by the four mutations provides a completed catalytic triad.

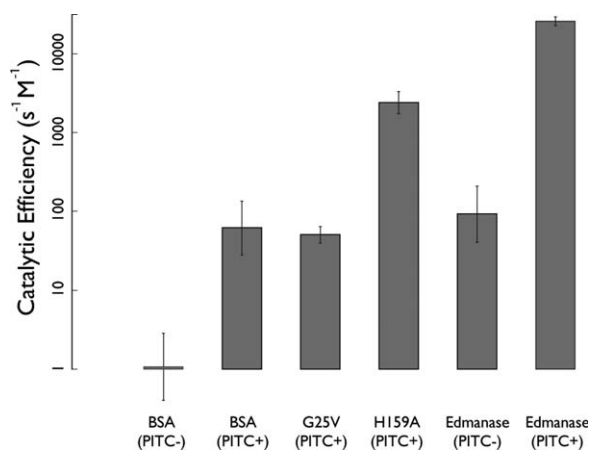


Figure 5. Catalytic efficiency of Edmanase and attenuation mutants. The catalytic efficiencies (k_{cat}/K_M) of the negative controls (BSA) and point mutants are compared to that of Edmanase. The substrate in each case was Ala-AMC, either untreated (PITC-) or derivatized with PITC (PITC+). BSA indicates values measured when Edmanase was replaced with BSA. The point mutants G25V and H159A serve to verify the intended catalytic mechanism of Edmanase by excluding the nucleophilic sulfur from the active site, and by removing the putative general base of the catalytic triad, respectively.

Table I. Kinetic Parameters of Edmanase for PTC-Ala-AMC Substrate Compared to Attenuation Mutants

Variant	k_{cat} (s^{-1})	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ ($\text{s}^{-1} \text{M}^{-1}$)	$k_{\text{cat}}(\text{Edmanase})/k_{\text{cat}}(\text{Mutant})$
Edmanase	0.55 (± 0.013)	21.3 (± 2.7)	2.6×10^4	1.0
G25V	0.064 (± 0.0034)	1270 (± 292)	50.4	0.116
H159A	0.24 (± 0.069)	96.1 (± 14.6)	2.4×10^3	0.436
BSA (PITC-)	0.0018 (± 0.0015)	1684 (± 877)	1.07	0.0033
BSA (PITC+)	0.0671 (± 0.037)	1090 (± 606)	61.5	0.122
Edmanase (PITC-)	0.0824 (± 0.058)	889 (± 367)	92.2	0.149

alanine to $\sim 10^1 \text{ s}^{-1} \text{ M}^{-1}$ for proline (Table II). The K_{M} values for different amino acids increase with the number of rotatable bonds in the side chain. This suggests that loss of side chain entropy upon complex formation may play a role in determining the catalytic efficiency. However, it is notable that Edmanase acts as an enzyme for all amino acid-AMC substrates tested.

Steric occlusion of the active site abrogates activity

We characterized several point mutants of Edmanase to determine if the substrate-assisted catalytic mechanism functioned as designed. We replaced G25 (formerly the position of the catalytic cysteine of cruzain) with a valine. The valine side-chain was expected to block access to the binding pocket, thus removing the potential for the remaining members

of the catalytic triad to contribute to activation of the sulfur nucleophile. Because of the substrate-assisted catalysis mechanism, we expected that occlusion of the catalytic site would lead to a near complete attenuation of Edmanase activity. The mutant was expressed and characterized using the same fluorescence based assay. The results demonstrate that only trace levels of activity remain, comparable to the BSA negative control, and that the attenuation of catalysis resulting from this substitution is ~ 500 -fold (Fig. 5, Table I).

Removal of the putative active site base attenuates activity

Our approach to design by substrate-assisted catalysis utilizes a nucleophile on the substrate molecule, but also requires a basic residue to enhance proton transfer from the substrate.⁵ This has been shown

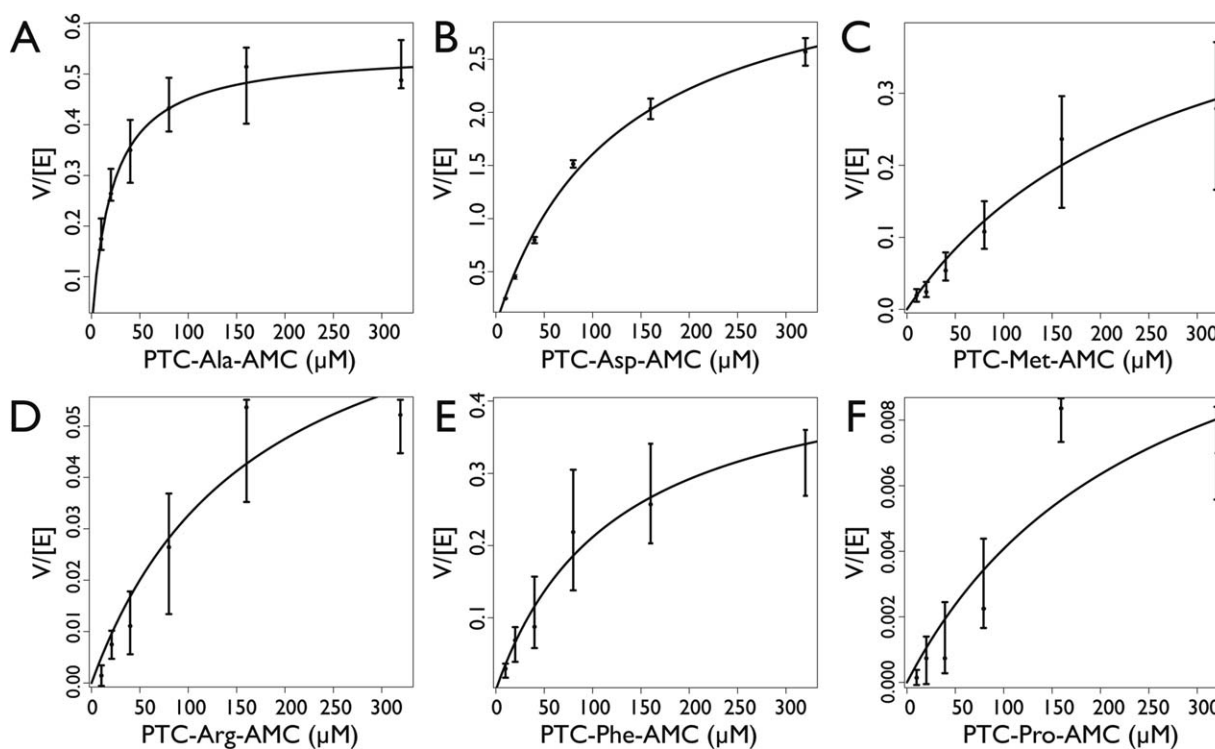


Figure 6. Edmanase exhibits limited specificity for the P1 position. Michaelis-Menten curves for the Edmanase enzyme for (A) PTC-Ala-AMC, (B) PTC-Asp-AMC, (C) PTC-Met-AMC, (D) PTC-Phe-AMC, (E) PTC-Arg-AMC, and (F) PTC-Pro-AMC are shown. These six amino acids represent a broad sampling of physicochemical characteristics of the canonical amino acid side chains. Edmanase cleaves each of the PITC-derivatized PTC-Xaa-AMC substrates with varying degrees of efficiency, with PTC-Asp-AMC being the most efficient and PTC-Pro-AMC being the least efficient.

Table II. Kinetic Parameters of Edmanase for PTC-Xaa-AMC Substrates

Substrate	k_{cat} (s^{-1})	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ ($\text{s}^{-1}\text{M}^{-1}$)
PTC-Ala-AMC	0.55 (± 0.013)	21.3 (± 2.7)	2.6×10^4
PTC-Asp-AMC	3.6 (± 0.41)	124.5 (± 35.0)	2.9×10^4
PTC-Phe-AMC	0.47 (± 0.060)	122.8 (± 29.8)	3.8×10^3
PTC-Met-AMC	0.54 (± 0.083)	271.8 (± 67.6)	2.0×10^3
PTC-Pro-AMC	0.0014 (± 0.0011)	252.0 (± 184.1)	5.7×10^1
PTC-Arg-AMC	0.087 (± 0.017)	167.8 (± 43.8)	5.2×10^2

to be required in other proteases that utilize a structurally similar catalytic triad.^{20–22} In Edmanase, this function is predicted to be fulfilled by His159. We therefore mutated this histidine to alanine, expecting a significant reduction in the catalytic activity. Results from the fluorescence activity assay are shown in Figure 5. The catalytic efficiency is reduced, although not as drastically as for the G25V mutant.

Catalysis is inhibited with the addition of a product analogue

The two mutations described above provide evidence that the mechanism of catalysis for Edmanase is the substrate-assisted mechanism that was intended. To provide further support, we attempted to identify an inhibitor for Edmanase. Searching through commercially available compounds yielded three potential molecules with elaborated thiazole rings that structurally mimic our product analog. We characterized the activity of Edmanase upon PTC-Ala-AMC with increasing concentrations of inhibitors. One of the three candidates, 1-(2-anilino-5-methyl-1,3-thiazol-4-yl)-ethanone showed inhibition of Edmanase. The

accumulation of cleaved substrate as a function of time for different concentrations of inhibitor are shown in Figure 7(A), and the inhibitory curve in Figure 7(B). The IC₅₀ is high (1.14 mM), indicating that 1-(2-anilino-5-methyl-1,3-thiazol-4-yl)-ethanone is only a weak inhibitor. Nevertheless, the fact that a structural mimetic for our putative product analog inhibits the enzyme provides further evidence that we have realized our desired substrate-assisted mechanism.

Discussion

We have succeeded in our goal of engineering an enzymatic activity in a neutral, aqueous environment to accomplish the cleavage step of the Edman degradation, a reaction that is normally carried out in harsh, acidic conditions. The ability to conduct controlled, step-wise degradation of peptides in biocompatible conditions allows for the inclusion of other biomolecules in assays to detect amino acid and peptide identity. Next-generation DNA sequencing strategies rely upon sensitive detection equipment that can visualize either single molecules or single PCR-amplified “clusters,” often using immobilization surfaces with excellent signal-to-noise properties. We expect that an enzymatic route to step-wise peptide degradation will prove compatible with these types of systems as well.

Our desire to catalyze the Edman cleavage step without altering its mechanism led us to pursue a substrate-assisted strategy. Several lines of evidence suggest that our Edmanase carries out the cleavage reaction using the mechanism that we intended.

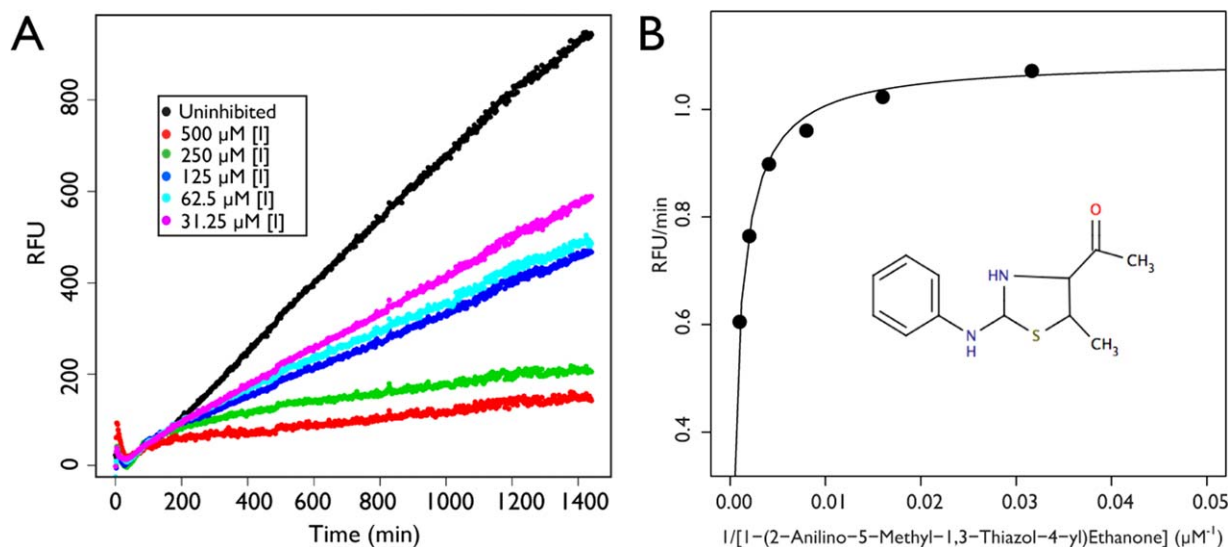


Figure 7. Inhibition of the Edmanase enzyme by a transition state analogue. (A) A series of fluorescence accumulation activity assays for the designed Edmanase enzyme in the presence of increasing inhibitor concentration are shown. PTC-derivatized Ala-AMC small molecules were used as the substrate for all assays. The enzyme is inhibited in a dose-dependent manner. (B). An inhibition curve fit to the data in panel A shows saturable inhibition with an IC₅₀ of 1.1 mM. A schematic of the structure of the inhibitor is shown in the inset.

First, release of the fluorescent AMC group is greatly increased when the fluorescent substrate is coupled with PITC. Second, Edmanase mutants in which either the putative catalytic site is sterically blocked, or in which the base of the catalytic triad is lost exhibit a drastic drop in catalysis. Finally, we were able to identify a weak competitive inhibitor for the enzyme by searching for a small molecule with structural similarity to the product analog that we used in our computational modeling.

We sought to engineer an Edmanase that could remove any N-terminal amino acid that has been derivatized with PITC. Towards this goal we selected as a design scaffold a protease with little selectivity at the corresponding P1 position. The model of the protein with the docked small molecule product analog confirmed that the methyl group corresponding to the side chain of an alanine residue was directed away from the enzyme and into solvent, limiting the potential for interactions with the protein. The judicious choice of such a scaffold allowed us to focus our computer-aided design on attempting to introduce favorable interactions with the phenyl ring of the PTC moiety, a constant for all substrates, and not on interactions with the P1 side chain.

While our engineered Edmanase exhibits activity against all of the substrates tested, including acidic, basic, small, large, and aromatic amino acids at the P1 position, we observed differences in the kinetic parameters for each. The lowest catalytic efficiency was observed for the PTC-Pro-AMC substrate; as an imino acid, proline is chemically different from other residues, and is also known to exhibit poor cleavage efficiency with the conventional Edman degradation protocol.²³ There are several potential solutions to the “proline problem.” Preliminary quantum chemical calculations suggest that the phenyl ring of the PTC-Pro conjugate adopts a rotated geometry relative to the PTC-Ala model used for computational design (data not shown). It may be possible to engineer a proline-specific Edmanase and use a combination of enzymes to cleave a mixed population of N-terminal amino acids. Alternatively, a different scaffold may be more appropriate for engineering a proline-specific Edmanase. Nature has evolved a variety of enzymes to cleave peptide bonds involving proline, such as prolyl aminopeptidase and Xaa-Pro dipeptidyl-peptidase. One of these enzymes may serve as a better starting point for design of an activity to remove PTC-Pro moieties from peptides.

Although our goal is to apply Edmanase to the cleavage of immobilized peptides, we have only demonstrated its activity towards PTC-Xaa-AMC substrates. We are optimistic that the enzyme will also catalyze the removal of amino acids from peptides because that activity is present in cruzain. However, sequence-specific complications within the context of peptide substrates may give rise to variability in

steric hindrance or catalytic efficiency. Ultimately, the utility of Edmanase will need to be assessed in the experimental immobilization and detection contexts developed for high-throughput peptide analysis.^{24,25}

While Edmanase exhibits excellent catalytic efficiencies towards several substrates with different amino acids at the P1 position (10^3 – 10^4 $M^{-1} s^{-1}$), the rate accelerations for these substrates are less impressive. For instance, the ratio of k_{cat} values for Edmanase and BSA utilizing a PTC-Ala-AMC substrate differ by roughly an order of magnitude, even though the ratio of catalytic efficiencies for these proteins is $\sim 10^3$. We speculate that our attempts to re-engineer the cruzain protease to recognize the PTC moiety on the substrate have been successful. However, we have not characterized a cruzain mutant with only the catalytic cysteine mutation, and thus cannot isolate the effects, positive or otherwise, of the mutations intended to recognize the phenyl group. We have apparently retained only a fraction of the catalytic power of the active site. Indeed, while a mutation that occludes the active site reduces activity to baseline levels (G25V mutant in Table II), the loss of the putative base in the catalytic triad results in only a modest decrease in k_{cat} . This raises the possibility that Edmanase could be further improved by mutations that restore the potency of the active site, or that bind the transition state in an orientation better aligned with the remaining members of the catalytic triad.

We believe that future protein engineering efforts will share many similarities with the work presented here. Although recent studies have demonstrated that computational enzyme design can create *de novo* active sites out of protein clefts and pockets,^{26–28} in many cases this impressive ability is likely to be unnecessary. Our goal was to generate a needed enzymatic activity using already developed modeling capabilities. To maximize the chance of achieving this goal we sought to minimize the number of necessary mutations made to a wild-type protein. This entailed both careful selection of a design scaffold, such that as many desired features for redesign as possible were already present in the starting template, and also manual curation of proposed mutations suggested by the RosettaDesign program to include only those predicted to interact with groups present in the substrate (here, the phenyl ring in the PITC group). As we have demonstrated with our four-fold mutant, even a modest amount of protein engineering is sufficient to generate novel activities to order.

Materials and Methods

Scaffold selection

Selection of the cysteine protease, cruzain (PDB: 1AIM¹⁰), as a design scaffold was done by manually

searching the PDB for a crystal structure of a protease meeting several criteria including: (1) Presence of a substrate with a terminal aromatic group, (2) utilization of a catalytic cysteine, (3) an active site in which the amino-acid side chain would be solvent exposed and not involved in substrate recognition, and (4) an active site in which a peptide substrate with a phenylalanine-like moiety could be accommodated. The search returned only this single protein scaffold as a potential hit.

Chemical modeling

An analogue for the immediate post-cleavage product of an alanine residue was modeled using quantum simulation in the GAMESS package,¹² and energy minimized under the 6-31G basis set. The catalytic cysteine in cruzain was mutated *in silico* to a glycine residue to accommodate the product analogue. The analogue was then docked into the active site of the cruzain scaffold by constraining the sulfur in the small molecule to overlap with the former position of the sulfur of the catalytic cysteine. The position and orientation of the modeled substrate within the active site were optimized by energy minimization using the Rosetta full atom energy function augmented by the harmonic constraint upon the small molecule sulfur atom.

Computer-aided design

Mutations to accommodate the PITC-modified substrate were chosen manually after computational saturation mutagenesis of active site residues. First, residues that made contact with the substrate were chosen by visual inspection. These residues were computationally mutated to reasonable replacements that were chosen based on well-known, native-like interaction motifs. In particular, the ring-to-edge motif between Tyr157 and the product, and the thiol-edge interaction between Cys133 and the product are known to be preferred packing arrangements.^{16,19} Rosetta energies for the individual and combined mutations were tabulated. The top mutations were visually inspected, and the 4-fold mutant was chosen because of the presence of the canonical interactions and the well-satisfied hydrogen bonding.

Substrates and inhibitors

Single amino acid, amido-methylcoumarin (Xaa-AMC) containing compounds were obtained from BAChem (Bubendorf, Switzerland). These included Arg-AMC, Asn-AMC, Phe-AMC, Met-AMC, Ala-AMC, and Pro-AMC. Phenylisothiocyanate (PITC) was purchased from Thermo-scientific and coupled to the N-terminus of each substrate by incubating for 20 min at 50° C in a 100 μ L solution of acetonitrile:pyridine:water (10:5:3) with 5 μ L of PITC. The derivatized substrate was then dried by rotary evap-

oration and resuspended in 250 μ L of 1x PBS. Inhibitor compound, 1-(2-anilino-5-methyl-1,3-thiazol-4-yl)-ethanone, was obtained from Sigma-Aldrich (St. Louis, MO).

Enzyme purification

A synthetic gene encoding the engineered enzyme was purchased from GenScript (Piscataway, NJ) and cloned into a pET42(a) (Novagen) expression vector between the NdeI and XhoI sites, then transformed into *E. coli*, BL-21(DE3) chemically competent cells. Protein was over-expressed following Studier's auto-induction protocol.²⁹ The protein was expressed at high levels and found in inclusion bodies. Bacterial cells were harvested by centrifugation of the cell culture at 5000 rpm and 4° for 10 min. Cells were then resuspended in 1x PBS pH 7.4 with 10% glycerol and 6M guanidine hydrochloride, pH 7.4. Cells were then lysed by sonication on ice. The cell lysate was centrifuged at 18,000 rpm, 4° for 20 min. The supernatant was then filtered through a 0.22 μ m cellulose acetate filter. The filtered lysate was loaded onto a 5 mL HisTrap (Ni-NTA) column and washed with five column volumes of binding buffer (50 mM Tris-HCl, 150 mM NaCl, 6M guanidine hydrochloride, 25 mM imidazole). Bound protein was then eluted in 50 mM Tris-HCl, 150 mM NaCl, 6M guanidine hydrochloride, 500 mM imidazole. Purified fractions were then refolded by successive, overnight dialyses into 1x PBS containing 5, 3, 1, 0.5, and 0M guanidine hydrochloride. Purified fractions were prepared for SDS-PAGE analysis by mixing two parts sample with one part 4x loading dye. Samples were analyzed on 16% SDS-PAGE precast gels, and visualized by Coomassie staining. Protein concentration was determined using the calculated molar extinction coefficient (A280)³⁰ and measured on an ND-8000 spectrophotometer (Thermo Fisher Scientific).

Activity measurements

All kinetic measurements were performed in a 96-well black Corning plate on a BioTek Synergy2 plate reader at 30°C. Reactions were started by addition of 5–20 μ L of purified enzyme to 150 μ L of 0.1–10 mM substrate solution. Final enzyme concentration was between 1 and 100 nM. The ratio between maximum substrate concentration and enzyme concentration was equal in all experiments. Fluorescence of the cleaved product was measured by excitation at 370 nm (30 s intervals for 1–10 h) and monitoring emission at 460 nm. Fluorescence accumulation was monitored every 30 s and relative fluorescence units were converted to rates of substrate cleavage by calibration with a free AMC standard curve (Sigma Aldrich, MO). Reaction rates at steady state were calculated from the slope of the fluorescence time courses by linear regression of initial velocities, and kinetic parameters were calculated assuming

Michaelis-Menten kinetics, $v = V_{\max}[S]/([S] + K_M)$ by nonlinear regression in the R statistical software package. For inhibition experiments, assays were conducted with 5 μM substrate, 100 nM enzyme, and 500–31.25 μM 1-(2-anilino-5-methyl-1,3-thiazol-4-yl)-ethanone. Reaction velocity was plotted against the inverse of inhibitor concentration, and fit by nonlinear least squares to determine the IC50.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. *Nat Rev Genet* 5:335–344.
- Edman P (1949) A method for the determination of amino acid sequence in peptides. *Arch Biochem* 22:475.
- Niall HD (1973) Automated Edman degradation: the protein sequenator. *Methods Enzymol* 27:942–1010.
- Dall'Acqua W, Carter P (2000) Substrate-assisted catalysis: molecular basis and biological significance. *Protein Sci* 9:1–9.
- Perona JJ, Craik CS (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci* 4:337–360.
- Carter P, Wells JA (1987) Engineering enzyme specificity by “substrate-assisted catalysis.” *Science* 237:394–399.
- Corey DR, Willett WS, Coombs GS, Craik CS (1995) Trypsin specificity increased through substrate-assisted catalysis. *Biochemistry* 34:11521–11527.
- Dall'Acqua W, Halin C, Rodrigues ML, Carter P (1999) Elastase substrate specificity tailored through substrate-assisted catalysis and phage display. *Protein Eng* 12:981–987.
- Berman H, Henrick K, Nakamura H, Markley JL (2006) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 35:D301.
- Gillmor SA, Craik CS, Fletterick RJ (1997) Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci* 6:1603–1611.
- Lecaille F, Authie E, Moreau T, Serveau C, Gauthier F, Lalmanach G (2001) Subsite specificity of trypanosomal cathepsin L-like cysteine proteases. *Eur J Biochem* 268:2733–2741.
- Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S (1993) General atomic and molecular electronic structure system. *J Comput Chem* 14:1347–1363.
- Pillay D, Boulangé AF, Coetzer THT (2010) Expression, purification and characterization of two variant cysteine peptidases from *Trypanosoma congolense* with active site substitutions. *Protein Express Purif* 74:264–271.
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popović Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 487:545–574.
- Borgo B, Havranek JJ (2014) Motif-directed redesign of enzyme specificity. *Protein Sci* 23:312–320.
- Burley SK, Petsko GA (1985) Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* 229:23–28.
- Serrano L, Bycroft M, Fersht AR (1991) Aromatic-aromatic interactions and protein stability: investigation by double-mutant cycles. *J Mol Biol* 218:465–475.
- Mitchell JBO, Nandi CL, McDonald IK, Thornton JM, Price SL (1994) Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? *J Mol Biol* 239:315–331.
- Reid KSC, Lindley PF, Thornton JM (1985) Sulphur-aromatic interactions in proteins. *FEBS Lett* 190:209–213.
- Barrett AJ, Rawlings ND (1996) Families and clans of cysteine peptidases. *Perspect Drug Discov Des* 6:1–11.
- Page MJ, Di Cera E (2008) Serine peptidases: classification, structure and function. *Cell Mol Life Sci* 65:1220–1236.
- Polgar L (2005) The catalytic triad of serine peptidases. *Cell Mol Life Sci* 62:2161–2172.
- Hermodson MA, Ericsson LH, Titani K, Neurath H, Walsh KA (1972) Application of sequenator analyses to the study of proteins. *Biochemistry* 11:4493–4502.
- Tessler LA, Donahoe CD, Garcia DJ, Jun YS, Elbert DL, Mitra RD (2011) Nanogel surface coatings for improved single-molecule imaging substrates. *J R Soc Interface* 8:1400–1408.
- Tessler LA, Mitra RD (2011) Sensitive single-molecule protein quantification and protein complex detection in a microarray format. *Proteomics* 11:4731–4735.
- Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retro-aldol enzymes. *Science* 319:1387–1391.
- Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453:190–195.
- Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329:309–313.
- Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Express Purif* 41:207–234.
- Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182:319–326.