This is the published version:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30038994

Reproduced with the kind permissions of the copyright owner.

Copyright : 2003, Elsevier (Open Access journal)
A Despecialization Step Underlying Evolution of a Family of Serine Proteases

Merridee A. Wouters,1 Ke Liu,1,2 Peter Riek, and Ahsan Husain1,*
Victor Chang Cardiac Research Institute
Sydney, New South Wales 2010
Australia

Summary

In the trypsin superfamily of serine proteases, non-trypsin-like primary specificities have arisen in only two monophyletic descendent subbranches. We have recreated an ancestor to one of these subbranches (granzyme) using phylogenetic inference, gene synthesis, and protein expression. This ancestor has two unusual properties. First, it has broad primary specificity encompassing the entire repertoire of novel primary specificities found in its descendents. Second, unlike extant members that have narrow primary specificities, the ancestor exhibits tolerance to mutational changes in primary specificity-conferring residues—that is, structural plasticity. Molecular modeling and mutagenesis studies indicate that these unusual properties are due to a particularly wide substrate binding pocket. These two crucial properties of the ancestor not only distinguish it from its extant descendents but also from the trypsin-like proteases that preceded it. This indicates that a despecialization step, characterized by broad specificity and structural plasticity, underlies evolution of new primary specificities in this protease superfamily.

Introduction

Serine proteases of the trypsin superfamily are broadly classified by their primary specificity. This specificity is based on the recognition of the substrate residue immediately N-terminal to the scissile bond (i.e., the substrate P1 residue). A comparison of crystal structures within this superfamily has led to a theory that 3 residues of the S1 substrate binding pocket confer primary specificity. These specificity-conferring residues are located at positions 189, 216, and 226 (chymotrypsin numbering) (Birktoft et al., 1970; Sloton and Hartley, 1970; Perona and Craik, 1997). For example, in trypsin, the aspartic acid in the D189-G216-G226 (DGG) triplet confers specificity for basic residues. In chymotrypsin, the small serine in its SGG triplet confers specificity for hydrophobics, particularly aromatics. For both of these enzymes, the glycines of the triplet located on the wall of the pocket allow access to the bulky residues preferred by these enzymes. In elastase these glycines are replaced with amino acids containing branched side chains (GVT) that restrict specificity to small aliphatics. Mutagenesis experiments have indicated the importance of these specificity-conferring residues. However, attempts to interchange primary specificity between superfamilies members through appropriate chimeric replacement of these residues virtually abolish catalytic activity (Venkeki et al., 1996; Hung and Hedstrom, 1998; Caputo et al., 1999; Solivan et al., 2002). The evolution of functional diversity in this superfamily has remained elusive because it is difficult to see how evolution could take place through intermediates that are functionally compromised (Perona and Craik, 1997).

The most plausible theory to explain the seeming paradox is that the ancestral protease of this superfamily was resistant to inactivation following a mutational change of its specificity-conferring S1 pocket residues because of structural plasticity of this S1 subsite. Such structural plasticity can be demonstrated for the S1 pocket of an invertebrate and a prokaryotic serine protease (Bone et al., 1989; Tsu et al., 1997). However, these proteases have broad multivalent primary specificities. This suggests that, in vertebrate evolution, primary specificities were progressively narrowed.

A group of vertebrate serine protease families (including enzymes of the kallikrein, granzyme, and complement factor D clusters) with a common ancestor mainly have narrow trypsin-like primary specificities. Surprisingly, several novel narrow primary specificities are seen in the recently evolved granzyme B/chymase subbranch of the granzyme family. This recent evolutionary event raises the question: how did novel narrow primary specificities evolve from an ancestor that itself had already acquired a narrow trypsin-like primary specificity?

We have used phylogenetic inference to predict the primary structure of the ancestor of a group of serine proteases comprising the granzyme B, α- and β-chymase, and cathepsin G protease clusters. These members of the trypsin superfamily are involved in cell-mediated immunity and will be referred to hereafter as immune defense proteases (IDPs). The extant IDPs have a number of non-trypsin-like primary specificities. An IDP, ancestral to members with non-trypsin-like primary specificities, was reconstructed by total gene synthesis and protein expression. We show that, unlike the extant members of this serine protease family, the recreated ancestor tolerates mutational changes in the specificity-conferring residues of the S1 pocket. However, the cost of this “structural and functional plasticity” is a relaxation of the presumed narrow trypsin-like primary specificity of its immediate predecessor—that is, the process appeared to reverse evolution. These findings give credence to the idea that in the trypsin superfamily of serine proteases the progenitor was a broad specificity protease with a plastic S1 substrate binding pocket.

Results

Chymase/Granzyme B Ancestor Has High Catalytic Efficiency and Multivalent Primary Specificity

Pauling and Zuckerkandl (1963) first suggested use of phylogenetic inference in predicting ancestral sequences.

*Correspondence: ahusain@physiology.uab.edu
1 These authors contributed equally to this work.
2 Present address: University of Alabama at Birmingham Physiology and Biophysics, 1530 3rd Avenue South - MCLM 968, Birmingham, Alabama 35294.
Figure 1. Primary Structure Prediction of Stemzyme-IDP-H9252

(A) Phylogenetic tree of the IDPs, as derived from parsimony analysis of 56 sequences using PAUP (Swofford, 1992). Stemzyme-IDP-H9252 is the ancestor that was reconstructed. In humans, descendents of node H9252 are located on chromosome 14, node H9253 on chromosome 19, and granzymes A and K on chromosome 5. Gene duplication after node H9252 resulted in two major branches of descendents: the granzymes/cathepsin G and the chymases (c). A recently sequenced platypus granzyme (Poorafshar et al., 2000) is basal to the granzyme branch, suggesting the enzyme created existed before the divergence of the earliest mammals approximately 170 million years ago (Killian et al., 2001; Wyss, 2001). Node α must be at least 450 million years old based on the presence of a trout ortholog of complement factor D (Hajnik et al., 1998; Kumar and Hedges, 1998). The trees were rooted using two closely related granzyme A sequences as an outgroup. Bootstrap percentages for 1000 replications are shown on each supported branch. Values below 70 are omitted for clarity. Common names of extant IDPs (Barrett et al., 1998) are shown in parentheses. GRAM, Granzyme M (Met-ase) specific for P1 methonine and leucine (Smyth et al., 1992); CAP7, CAPT, azuricidin/cap37/heparin binding protein, inactive; ELNE, neutrophil elastase, PRNS myeloblastin elastase-like specificity; CFAD, complement factor D, specific for P1 arginine; C1G, cathepsin G is specific for P1 phenylalanine and lysine as is DDN1 (duodenase); and MCTX, sheep sequence MCT1. The designation MCT generally refers to proteases found in mast cells. GRAx and GRLn designate granzymes, enzymes found in granules of cytotoxic lymphocytes. Of those characterized, granzyme B is specific for P1 aspartic acid (Poe et al., 1991), and granzymes A and K are specific for P1 arginine and lysine (Barrett et al., 1998). NKPn have been isolated from natural killer cells. Species codes that may require explanation are MERUN, Meriones unguiculatus, gerbil; MESAU, Mesocricetus auratus, Golden hamster; PAPHA, Papio hamadryas, baboon; MACFA, Macaca fascicularis, macaque. Specificity-conferring triplets for extant proteases and several evolutionary nodes are shown in green.

Figure 1. Primary Structure Prediction of Stemzyme-IDP-β

(B) Phylogenetic tree of the IDPs, as derived from parsimony analysis of 56 sequences using PAUP (Swofford, 1992). Stemzyme-IDP-β at node β is the ancestor that was reconstructed. In humans, descendents of node β are located on chromosome 14, node γ on chromosome 19, and granzymes A and K on chromosome 5. Gene duplication after node β resulted in two major branches of descendents: the granzymes/cathepsin G (β) and the chymases (ε). A recently sequenced platypus granzyme (Poorafshar et al., 2000) is basal to the granzyme branch, suggesting the enzyme created existed before the divergence of the earliest mammals approximately 170 million years ago (Killian et al., 2001; Wyss, 2001). Node α must be at least 450 million years old based on the presence of a trout ortholog of complement factor D (Hajnik et al., 1998; Kumar and Hedges, 1998). The trees were rooted using two closely related granzyme A sequences as an outgroup. Bootstrap percentages for 1000 replications are shown on each supported branch. Values below 70 are omitted for clarity. Common names of extant IDPs (Barrett et al., 1998) are shown in parentheses. GRAM, Granzyme M (Met-ase) specific for P1 methonine and leucine (Smyth et al., 1992); CAP7, azuricidin/cap37/heparin binding protein, inactive; ELNE, neutrophil elastase, PRNS myeloblastin elastase-like specificity; CFAD, complement factor D, specific for P1 arginine; C1G, cathepsin G is specific for P1 phenylalanine and lysine as is DDN1 (duodenase); and sheep sequence MCT1. The designation MCT generally refers to proteases found in mast cells. GRAx and GRLn designate granzymes, enzymes found in granules of cytotoxic lymphocytes. Of those characterized, granzyme B is specific for P1 aspartic acid (Poe et al., 1991), and granzymes A and K are specific for P1 arginine and lysine (Barrett et al., 1998). NKPn have been isolated from natural killer cells. Species codes that may require explanation are MERUN, Meriones unguiculatus, gerbil; MESAU, Mesocricetus auratus, Golden hamster; PAPHA, Papio hamadryas, baboon; MACFA, Macaca fascicularis, macaque. Specificity-conferring triplets for extant proteases and several evolutionary nodes are shown in green.
More recently this technique has been used to predict primary structures of ancient proteins, and the function of these ancient proteins has been tested following expression of the recreated gene (Malcolm et al., 1990; Stackhouse et al., 1990; Jermann et al., 1995; Chandrasekharan et al., 1996; Chang et al., 2002). We have previously applied these methodologies to recreate the molecular history of the chymase cluster of serine proteases (Chandrasekharan et al., 1996). In the current study, similar approaches were used to recreate a more distant IDP ancestor.

We inferred ancestral sequences of nodes α, β, and γ from parsimony analysis of a multiple sequence alignment of IDPs (Figure 1A). Of these three sequences, the β node could be predicted with the highest degree of certainty (Figure 1B). Stemzyme-IDP-β exhibits 50% to 69% sequence identity with extant β-IDPs shown in Figure 1A. Among these β-IDPs, stemzyme-IDP-β shares the least identity with sequences of the mouse granzyme D cluster, but the highest identity with nonrodent α-chymases, suggesting that the original function of the enzyme may have been α-chymase-like (Chandrasekharan et al., 1996). Stemzyme-IDP-β exhibits 36% to 48% sequence identity with extant non-β-IDPs shown in Figure 1A. With respect to these non-β-IDPs, stemzyme-IDP-β shares the least identity with members of the cap37/elastase cluster and highest identity with granzymes A and K. The S, pocket triplet in stemzyme-IDP-β is AGG (Figure 1B), 1 amino acid different from the S, pocket triplet DGG predicted for the α-node. A synthetic gene encoding the enzyme was constructed and the recombinant protein expressed.

For kinetic studies, the octapeptide angiotensin II (Ang II, DRVYIHPF) and its position 4 analogs were used as substrates because they are substrates for several β-IDPs (Barrett et al., 1998). For example, many chymases hydrolyze the Y4-I5 bond in Ang II (Barrett et al., 1998). In addition to the key P1-S interaction, multiple synergistic interactions between the enzyme’s extended substrate binding site (i.e., substrate binding sites on the enzyme other than those that are responsible for determining primary specificity) and Ang II allow an equivalent registration of the substrate. That is, the extended substrate binding site facilitates the position 4 amino acid in Ang II in being recognized by the enzyme as the sole P1 residue. The consequence of these extended substrate binding site interactions with Ang II is that when this position 4 P1 residue is suboptimal, the Ang II analog is either poorly cleaved at the 4–5 bond or not cleaved at all (Sanker et al., 1997; K.L. and A.H., unpublished data) rather than being cleaved at another site on the substrate. Equivalent registration of Ang II and its position 4 analogs is also seen with stemzyme-IDP-β. This is evident from studies with these substrates, which revealed that, although hydrolysis efficiency differed, cleavage always occurred at the 4–5 bond (Figures 2A–2L).

Kinetic studies with stemzyme-IDP-β demonstrate that it has a broad P1 specificity for cleavage of the 4–5 bond in Ang II analogs and displays high turnover numbers (kcat = 20 s⁻¹) for optimal P1 residues (Figures 2A–2L and Table 1). Components of stemzyme-IDP-β’s broad specificity include chymotrypsin-like, elastase-like, Glu/Asp-ase, and Met-ase-like activities. Most of these activities are found individually in descendant extant proteases, with chymases being relatively specific for tyrosine and phenylalanine (Powers et al., 1985) and granzyme B for aspartic acid (Poe et al., 1991). The dual phenylalanine/lysine P1 specificity of cathepsin G (Rehault et al., 1999), where the specificity ratio between substrates with phenylalanine as compared to a lysine at the P1 site is 12:5:1 to 20:1, is observed in stemzyme-IDP-β where the corresponding activity ratio is 100:1. Stemzyme-IDP-β also has elastase-like activity, which appears to be a property of rodent α-chymases (Solivan et al., 2002). Stemzyme-IDP-β also shows specificity for methionine, which is not found in β-IDPs but occurs in the Met-ase granzyme M (Barrett et al., 1998), a descendant of the ancestor found at node γ (Figure 1A). This broad primary specificity of the β-IDP ancestor does not appear to be a feature of any extant IDP.

**Mutation of S. Specificity-Conferring Residues in Stemzyme-IDP-β Does Not Abrogate Catalytic Activity**

Point mutations in the S, binding site triplet of extant serine proteases of the pancreatic and IDP group abrogate protease activity (Venekila et al., 1996; Hung and Hedstrom, 1998; Caputo et al., 1999; Solivan et al., 2002). We examined the effect of mutation of key residues in the S, binding pocket on stemzyme-IDP-β catalytic activity and their effects on P1 specificity. In the S, pocket were mutated to the triplets found in the ancestor and the catalytic activity of the enzyme is restored (Figure 1A). We also introduced the chymotrypsin triplet SGG and two related permutations, AGA and DGA. All of these combinations are found in extant IDPs (with the exception of DGA).

The AGA—SGG mutant had enhanced activity toward nearly all P1 residues tested, with most improvement toward chymotrypsin-like substrates (i.e., those with aromatics, particularly tryptophan) (Figure 3A). AGA and
Figure 2. Position 4 Analogs of Ang II Are Cleaved by Stemzyme-IDP-β between the 4–5 Bond Despite Nonoptimal P1 Side Chains
(A–L) HPLC chromatograms of the cleavage by stemzyme-IDP-β of Ang II (DRVY↓IHPF) or its analogs containing at position 4 an alanine (A), arginine (B), aspartic acid (C), glutamine (D), glutamic acid (E), leucine (F), lysine (G), methionine (H), phenylalanine (I), tryptophan (J), tyrosine (K), or valine (L). The invariant product IHPF and the substrate-dependent N-terminal tetrapeptide products are indicated by arrows. In certain cases this latter product was too hydrophilic and eluted at a retention time associated with the injection artifact that occurred between 0.4 and 0.8 min; in this case its elution position is not shown. The retention times of the substrates and products were ascertained using pure synthetic standards. Wherever possible, stemzyme-IDP-β concentration and incubation time were adjusted to obtain approximately 50% cleavage of the substrate.
Table 1. Kinetic Constants for the Hydrolysis of the X4-I5 Bond in Ang II (Contains a Tyr at Position 4) and Its Analogs by Stemzyme- IDP-β

<table>
<thead>
<tr>
<th>Substrate</th>
<th>kcat/KM (mM)</th>
<th>kcat (s⁻¹)</th>
<th>kcat/KM (s⁻¹·mM⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ang II</td>
<td>0.022 ± 0.0015</td>
<td>18.4 ± 0.4</td>
<td>850</td>
</tr>
<tr>
<td>[Met]Ang II</td>
<td>0.076 ± 0.0023</td>
<td>26.6 ± 0.2</td>
<td>350</td>
</tr>
<tr>
<td>[Val]Ang II</td>
<td>0.065 ± 0.0055</td>
<td>17.5 ± 0.4</td>
<td>270</td>
</tr>
<tr>
<td>[Leu]Ang II</td>
<td>0.11 ± 0.0015</td>
<td>21.5 ± 0.2</td>
<td>200</td>
</tr>
<tr>
<td>[Ala]Ang II</td>
<td>0.13 ± 0.0074</td>
<td>23.6 ± 0.2</td>
<td>190</td>
</tr>
<tr>
<td>[Phe]Ang II</td>
<td>0.13 ± 0.014</td>
<td>23.7 ± 0.9</td>
<td>180</td>
</tr>
<tr>
<td>[Glu]Ang II</td>
<td>0.17 ± 0.0092</td>
<td>19.3 ± 0.1</td>
<td>111</td>
</tr>
<tr>
<td>[Trp]Ang II</td>
<td>0.12 ± 0.0015</td>
<td>11.7 ± 0.5</td>
<td>99</td>
</tr>
<tr>
<td>[Asp]Ang II</td>
<td>0.24 ± 0.029</td>
<td>15.3 ± 0.5</td>
<td>64</td>
</tr>
<tr>
<td>[Gln]Ang II</td>
<td>0.37 ± 0.01</td>
<td>21.8 ± 0.5</td>
<td>59</td>
</tr>
<tr>
<td>[Arg]Ang II</td>
<td>–</td>
<td>–</td>
<td>4.8</td>
</tr>
<tr>
<td>[Lys]Ang II</td>
<td>–</td>
<td>–</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Values are the mean ± SEM of three independent measurements. Where individual kcat and KM values could not be determined independently (indicated by a dash) due to weak binding and limited solubility of substrates, kcat/KM was determined from triplicate measurements of velocity (v) at low substrate concentration ([S] < < Kd), where v = (kcat/Kd)[E][S].

SGA are found in extant chymases, and these mutations improved relative specificity of stemzyme-IDP-β toward tyrosine. AGA, found in rat chymase 2, improved the relative specificity toward tyrosine by decreasing its specificity toward most other residues tested (Figure 3A); that is, the enzyme became more chymase-like by decreasing its neck of the pocket (by ~1 Å) in descendents of the ancestral chymase, had a deleterious effect on the specificity of all residues tested, with phenylalanine being the least affected (Figure 3A). This mutation phenotype is consistent with a chymase primary specificity.

Attempts to reconstitute trypsin-like activity in stemzyme-IDP-β by introducing the DGG triplet found in trypsin resulted in an active enzyme with chymotrypsin-like properties. Despite an improvement in the specificity toward lysine at the expense of other P1 residues, hydrolysis at the P1 lysine or arginine remained poor (Figure 3A). The DGA triplet produced an improvement in specificity toward P1 lysine and arginine along with an overall impairment in catalytic efficiencies (Figure 3A), but the mutant still retained better specificity for P1 phenylalanine and leucine substrates. From these collective studies we also note that the degree of bias toward a P1 specificity created by the introduction of any particular triplet in stemzyme-IDP-β is much smaller than that seen in an extant IDP with the same triplet—that is, the specificity of stemzyme-IDP-β cannot be substantially narrowed by simple changes in the specificity-determining residues of the S1 pocket (Figure 3B).

Structural Basis of the Multivalent Specificity of Stemzyme-IDP-β

We next explored the basis for the broad primary specificity of stemzyme-IDP-β using structure and sequence comparisons. Multiple alignments of ancestral sequences with those of extant IDPs indicate that a 3 residue deletion occurred in the loop joining the pocket walls containing residues 216 and 226 between the α and β nodes (Figure 4A). The deletion makes this loop shorter in stemzyme-IDP-β and its descendents than in most other serine proteases (Figure 4B). The shorter loop is accommodated by a cis-peptide bond between residues 224 and 225, specific to extant β-IDPs. These changes result in a remodeled S1 pocket that is narrower at the base than most members of the trypsin superfamily (Figure 4B). The 3 residue deletion includes C220 and also results in the absence of a C191-C220 linkage (sequences 4–7, Figure 4A) that is conserved in other disulphide-containing serine proteases of the trypsin superfamily. The lack of this disulphide bond broadens the neck of the pocket (by ~1 Å) in descendents of stemzyme-IDP-β (Remington et al., 1988; Hof et al., 1996; McGrath et al., 1997; Pletnev et al., 2000; Waugh et al., 2000). Sequence and structure comparisons suggest that the S1 triplet, the neck of the pocket, and conserved water molecules in the pocket (Sreenivasan and Axelsen, 1992) have been individually customized in at least three major clusters: the chymase cluster (α) and the two granzyme clusters descending from node δ. For example, in the cathepsin G/granzyme B cluster, the residue at position 226 has assumed the charge-charge-coupling function from that residue at position 189 in trypsin-like enzymes which usually subserves this function (Figure 4A). Molecular modeling suggests that the neck of the pocket is especially wide in stemzyme-IDP-β due to the retention of the ancestral C191, which lines the inner lip of the pocket (data not shown). This may contribute to broad specificity by allowing the docking of a broader range of substrate P1 residues. In cathepsin G/granzyme B and chymase clusters the neck is narrowed by the presence of a Phe or Tyr at position 191. In the Grl2 cluster, which has a highly unusual S1 triplet (ARR), it is customized to glycine.

Cysteine 191 to Phenylalanine Mutation Establishes Chymotrypsin-like Specificity in Stemzyme-IDP-β

To test the effect of pocket-neck geometry on substrate specificity, we introduced a phenylalanine in place of cysteine at position 191 in stemzyme-IDP-β; a phenylalanine occurs naturally at position 191 in many IDP-β descendents and is expected to narrow the neck of the S1 pocket. Importantly, because the specificity-conferring S1 triplet in stemzyme-IDP-β is the same as that found in some chymotrypsin-like enzymes, we expected the emergence of chymotrypsin-like specificity in stemzyme-IDP-β with an engineered, narrow-necked S1 pocket. The C191F mutation was generally deleterious for substrate hydrolysis but not uniformly so. Catalytic efficiencies for the hydrolysis of substrates with P1 residues favored by chymotrypsin-like enzymes (phenylalanine, tyrosine, tryptophan, methionine, and leucine) were, on average, ~8-fold higher (p < 0.005) than those for substrates with other P1 residues (alanine, glutamine, lysine, arginine, valine, aspartic acid, and glutamic acid) (Figure 5A). This is in marked contrast to wild-type stemzyme-IDP-β where this difference was not significant (Figure 5B). Thus, the C191F mutation in stemzyme-IDP-β made it more chymotrypsin-like and hence narrowed its primary specificity.

The prediction by molecular modeling that the neck...
Figure 3. Mutational Changes in Stemzyme-IDP-β Specificity-Conferring Triplet (AGG) Do Not Abrogate Activity but Also Do Not Substantially Narrow P₁ Specificity

(A) AGG→DGG, AGG→SGG, and AGG→AGA mutations in the stemzyme-IDP-β specificity-conferring triplet generally produced modest effects on the specificity constant ($k_{cat}/K_m$) for P₁ residues preferred by chymotrypsin-like enzymes (orange circles: P₁ phenylalanine, tyrosine, methionine, leucine, tryptophan), trypsin-like enzymes (dark blue circles: P₁ lysine, arginine), Asp/Glu-ases (red circles: P₁ aspartic acid, glutamic acid), elastases (green circles: P₁ alanine, valine), and other P₁ specificities (light blue circles: P₁ glutamine). However, the AGG→DGA and AGG→SGA mutations had a more substantial detrimental effect on these P₁ specificities. Each value represents the average of at least three independent $k_{cat}/K_m$ measurements.

(B) Stemzyme-IDP-β specificity-conferring triplet mutations in (A) that are found in chymotrypsin-like enzymes (SGG, AGA, and SGA) or other triplet combinations (DGG and DGA) did not substantively narrow the chymotrypsin-like P₁ specificity of stemzyme-IDP-β. Values are the means ±1 SEM (error bars).
Serine Protease Evolution

Figure 4. Changes in the S1 Pocket in Stemzyme-IDP-β, Its Descendents, and Related Serine Proteases

(A) Alignment of representative IDPs with other serine proteases of the trypsin superfamily in the region of the S1 pocket. The first 8 sequences are extant IDPs. Sequence 9 is stemzyme-IDP-β. Sequences 10–13, included for comparison, are diverse members of the S,A family (http://www.merops.ac.uk): pancreatic trypsin (TRY1) and chymotrypsin B (CTRB), and vitamin K-dependent serine proteases, factor X (FA10), and thrombin (THRB). The highly conserved disulphide (yellow) found almost universally in the S1A family was lost at node β, possibly concomitant with a deletion in the vicinity of C220. S1 specificity-conferring triplet residues are in red. The active site serine (S195) is in white on a red background. Arrows below the alignment indicate conserved β strands. Boxed sequence in GRAA_HUMAN represents the segment that was introduced in stemzyme-IDP-β in place of the boxed sequence in STEMZYME-IDP-β to create the chimera [GraA-loop]-stemzyme-IDP-β.

(B) Stereo view from the side of the S1 site of the stemzyme-IDP-β descendent cathepsin G with a P1 phenylalanine bound (magenta). Several elements of bovine trypsin structure are superimposed for comparison. The remodeled pocket wall of the β-IDPs (residues 221 to 225, orange loop) substantially narrows the pocket bottom compared to serine proteases that diverged prior to the β node, which have trypsin-like pocket walls (red loop). The C191-C220 disulphide in trypsin (in yellow) is at the neck of the pocket behind the P1 phenylalanine; this is lost in stemzyme-IDP-β and its descendents. Customization of the pocket particular to a subset of β-IDPs (Figure 1) includes transferal of the role of primary specificity-determinant from residue 189 to residue 226. In the β-IDP cathepsin G, the E226 side chain (orange) projects across the base of the pocket away from the viewer, while in trypsin, D189 (red) points from the back of the pocket wall toward the viewer. This modification effectively moves the carboxylic acid moiety 1–2 Å away from the relocated pocket wall in cathepsin G, enabling water to penetrate between the end group and the pocket wall.

of the stemzyme-IDP-β S1 pocket is narrowed by the C191F mutation (data not shown) is consistent with the finding of a >100-fold decrease in catalytic efficiency for the hydrolysis of a substrate containing a P1 valine (Figure 5), since valine with its short, bulky, β-branched side chain is expected to fit poorly in S1 pockets with narrow necks. On the other hand, hydrolysis of the substrate with a P1 alanine, where the side chain is short and unbranched, was much less affected (6-fold decrease) by this mutation (Figure 5).

Markedly detrimental effects of the C191F mutation were also observed on the hydrolysis of substrates with acidic side chains in the P1 position (30- to 74-fold decrease in hydrolysis of substrates with P1 aspartic or glutamic acid) (Figure 5). But the finding that the C191F mutation had much less effect (6.5-fold decrease) on the hydrolysis of a substrate containing P1 glutamine, a glutamic acid-isostere with an uncharged, polar side chain, indicates that the nonproductive interaction is more specifically due to the presence of a terminal negative charge. On the other hand, acidic P1 side chains may adopt a more energetically favorable conformation in the less-restricted S1 pocket of wild-type stemzyme-IDP-β.

For the hydrolysis of substrates with a basic P1 side chain, the overall catalytic specificity was low in stemzyme-IDP-β and remained low after the introduction of the C191F mutation. In lysine and arginine, the terminal charge on the side chain is positive; but here, the charge is located much further away from the α-carbon than in aspartic acid and glutamic acid and could indicate that in stemzyme-IDP-β the catalytic specificity for these P1
residues is more likely to be influenced by the geometry and nature of the S₁ pocket base than by its portal.

A Wider S₁ Pocket Base Better Accommodates Long Bulky P₁ Side Chains

Molecular modeling of the stemzyme-IDP-β S₁ pocket predicts that it is narrower at the base than most members of the trypsin superfamily (Figure 4B). The remodeling that occurs with the evolutionary transition from node a to β (Figure 1A) is strongly influenced by a 3 residue deletion in the base-loop of the S₁ pocket. The narrowing of the S₁ pocket base is expected to be particularly detrimental to the hydrolysis of substrates with long and bulky P₁ side chains, such as those of arginine and lysine, because of poor accommodation in a pocket with a narrow base.

To study the effect of a wider S₁ pocket base on hydrolysis of substrates with basic P₁ side chains, our initial objective was to reintroduce in the loop joining that of alanine, is expected to experience steric hindrance while docking with a narrow-necked S₁ pocket. Importantly, however, compared to both wild-type stemzyme-IDP-β and its C191F mutant, the main effect of introducing the human granzyme A base-loop in stemzyme-IDP-β was the marked improvement in catalytic efficiency for the hydrolyses of substrates with large and bulky P₁ side chains, such as, lysine, arginine, phenylalanine, leucine, and tryptophan, the exception being tyrosine (Figures 5 and 6).

The further introduction of the A177D mutation in [GraA-loop]-stemzyme-IDP-β is identical to that in stemzyme-IDP-β. A substrate with valine, but not alanine, in the P₁ position was particularly poorly hydrolyzed by [GraA-loop]-stemzyme-IDP-β (Figures 5 and 6). This provides kinetic evidence that the introduction of the granzyme A base-loop in stemzyme-IDP-β caused a narrowing of the S₁ pocket neck. The β-branched side chain of valine, unlike that of alanine, is expected to experience steric hindrance while docking with a narrow-necked S₁ pocket. Importantly, however, compared to both wild-type stemzyme-IDP-β and its C191F mutant, the main effect of introducing the human granzyme A base-loop in stemzyme-IDP-β was the marked improvement in catalytic efficiency for the hydrolyses of substrates with large and bulky P₁ side chains, such as, lysine, arginine, phenylalanine, leucine, and tryptophan, the exception being tyrosine (Figures 5 and 6).

The further introduction of the A177D mutation in [GraA-loop]-stemzyme-IDP-β changes the S₁ specificity-conferring triplet from AGG to DGG (found in trypsin and granzyme A). This produced a marked overall de-
increase in catalytic efficiency (data not shown). Even though the mutant enzyme remained essentially chy motrypsin-like, the trypsin-like phenotype improved by >8-fold with the introduction of the A177D mutation; the P1 arginine to P1 tyrosine ratio of $k_{cat}/K_m$ values was 1:177 for the wild-type stemzyme-IDP-$\beta$, 1:28 for its GraA-loop mutant, and 1:3.4 for its A177D/GraA-loop mutant.

These findings show that mutations that are likely to alter $S_1$ pocket geometry can narrow the broad primary specificity of stemzyme-IDP-$\beta$, perhaps better than those mutations that change the specificity-conferring triplet. These structural aspects of the $S_1$ pocket hence are an important part of the molecular basis of primary specificity in stemzyme-IDP-$\beta$. Moreover, it is particularly noteworthy that high catalytic efficiencies are retained, and in some cases even enhanced, in stemzyme-IDP-$\beta$ despite substantial modification of the $S_1$ pocket. This demonstrates the functional tolerance of stemzyme-IDP-$\beta$ to structural change, a feature which is not to our knowledge found in its extant descendents.

**Discussion**

The high catalytic efficiency of stemzyme-IDP-$\beta$ is remarkable because it suggests that this deep evolutionary reconstruction (stemzyme-IDP-$\beta$ is ~170 million years old) is accurate. The reconstruction of stemzyme-IDP-$\beta$ also represents the identification of an evolutionary missing link that gave rise to primary specificity diversification in serine proteases. Previous studies had alluded to a gradual narrowing of primary specificity, from the initial ancient serine protease with a broad primary specificity, as the mechanism by which diversity was achieved in this protease superfamily (Perona and Craik, 1997). Our findings support this view but suggest that once a narrow primary specificity had become established, the further generation of diversity required a reversion to the presumed original state. That is, specialization or evolution-in-reverse was required for further diversification.

In most extant serine proteases, even conservative mutational changes in the specificity-conferring triplet virtually abolish catalytic activity (Venekel et al., 1996; Hung and Hedstrom, 1998; Caputo et al., 1999; Solivan et al., 2002), and for this reason it has been suggested that evolutionary changes have customized the $S_1$ pocket. This customization not only optimizes the shape of the $S_1$ pocket to allow proper fitting of the P1 side chain but helps to position the triplet side chains in an orientation that allows this interaction to be productive (e.g., Graf et al., 1987). The inability to accurately interconvert primary specificities between superfamly members even after substantive chimeric replacements (Hung and Hedstrom, 1998; Hedstrom et al., 1994) further suggests that global (protein-wide) as well as local ($S_1$ pocket) changes are responsible for this customization. Protein-wide evolutionary adjustments in fine structure associated with this customization could explain why only two major monophyletic groups of serine proteases with new primary specificities have emerged on the trypsin-like phenotypic background over the last 500 million years; these are the pancreatic chymotrypsin/elastase branch, which emerged during invertebrate evolution, and the more recently evolved $\beta$-IDP branch. All of this makes the reconstruction of an active stemzyme-IDP-$\beta$ even more remarkable.

The changes that occurred in evolution between the IDP ancestor found at node $\alpha$ and its closest identifiable descendent found at node $\beta$ (Figure 1A) allowed diverse P1 side chain docking conformations without significantly compromising catalytic efficiency. No less important was the appearance of another phenotype that to our knowledge has not been seen in extant IDPs: functional tolerance to mutations in the specificity-conferring triplet. These phenotypes likely had two important consequences. First, the $\beta$-IDP ancestor became free to explore mutations in specificity- and $S_1$ pocket shape-determining residues without a heavy inactivation penalty. This rare transition followed by progressive recustomization eventually led to the creation of novel, narrow primary specificities. Second, the $\beta$-IDP ancestor had a broad range of novel primary specificities. In the evolu-
tion of the immune system, the appearance of stem-
zyme-IDP-β with its novel specificities must have been
highly advantageous since its extant daughters, which
have its individual P1 specificities, play key roles in host
defense (Heusel et al., 1994). Its ability to tolerate mutations
that specialize function without loss of catalytic activity may underlie its success.

Important clues about the influence of S1 pocket ge-
ometry in shaping a P1 specificity have come from stud-
ies of an invertebrate and a prokaryote serine protease
(Bone et al., 1989; Tsu et al., 1997). These proteases,
which have wide-necked S1 pockets, have broad multiva-
lent primary specificities and demonstrable structural
plasticity with respect to the specificity-conforming tripl-
et. It is self-evident that a greater variety of substrate P1,
side chains and specificity-conferring triplet side chains
can be accommodated in a larger binding pocket. Our
findings suggest that among the many evolutionary
changes that occurred in producing the stemzyme-IDP-β
phenotype was a key mutation that modified the S1
pocket of an ancient IDP with a trypsin-like phenotypic
background (found at node α [Figure 1A]). This mutation,
a 3 residue deletion in the loop joining the pocket walls
containing residues 216 and 226 (Figure 4A), simultane-
ously caused the pocket base to narrow and the neck
to widen. The wider neck broadened P1 specificity. Due
to poor fit, however, the narrow base appears to have
been particularly detrimental for accommodating the long
side chains of P1, arginine- and lysine-containing
substrates that are preferred by trypsin and could ex-
plain why a trypsin-like P1 specificity does not seem to
have reemerged in extant β-IDPs.

Previously it was thought that the role of the S1 pocket
in these serine proteases was primarily to facilitate the
initial docking of the substrate—that is, in the formation
of the Michaelis enzyme-substrate complex, which is
also known as the ground state. It is now appreciated
that the main role of the S1 pocket is in effecting a narrow
primary specificity in its ability to facilitate the chemi-
cal steps of the catalytic reaction (Hedstrom et al.,
1994)—that is, those enzyme-substrate interactions that
reduce the Gibb’s free energy (ΔG‡) required to form
the transition state. In keeping with this evolving con-
cept, we find that mutations that alter S1 pocket shape,
such as those that are likely to narrow the S1 pocket
neck in stemzyme-IDP-β or widen its base markedly
influenced the transition state (Figure 6) but generally
had much smaller effects on the ground state (data not
shown). In the evolution of extant IDPs, mutations that
changed S1 pocket shape could also have served to
reduce freedom of movement of the P1 side chain within
its pocket, thereby optimally positioning the vicinal scis-
sile bond. Such optimization is expected to increase the
rate of reaction but also necessarily makes it difficult to
change a particular P1 specificity by altering its S1 triplet;
this could explain why low plasticity is a recurrent fea-
ture, not only in extant IDPs but also in most proteases
of the trypsin superfamly.

Experimental Procedures

Ancestral Sequence Prediction

Fifty-six IDP protein sequences were aligned using ClustalW
(Thompson et al., 1994) and adjusted using Cameleon (V3.14a Oxford
Molecular). Sequence codes are from SWISS-PROT as shown in
Figure 1A except GenBank Sequences: GRAM_MOUSE (AAD51606)
ELNE_MOUSE (AAD60670), BPH_MOUSE (AAD58256), CFAM_TROUT
(CAC79927), MCT3_SHEEP (CA73859), MCT2_SHEEP (CA69327),
MCT1_MESAU (BA199932), MCT4_RAT (AAD82660), MCT9_RAT (AA865246),
GRAJ_RAT (AAC53168), GRAR_RAT (AC17930). A corresponding RNA alignment was also created. A phylogenetic tree was created over 1000 bootstrap replications based on a maximum parsimony analysis of the protein alignment using PAUP (Swofford, 1992). Multiple starting trees were generated using random addition. A maximum-likelihood tree, calculated using PUZZLE (Strimmer and vonHaseeler, 1996), exhibited a high level of congruence to the maximum parsimony tree. The maximum parsimony
and maximum likelihood trees for the granzyme subbranches
are identical. The differences in branching were few and largely
confined to some remaining genes in the subtree which includes
the elastases, Met-ases, and complement factor D. The node recon-
structed is in the part of the tree with a high degree of congruence
between the two methods. For node βl, PAUP predicted 72% of the
227 residue ancestral sequence unambiguously. Assignments for
the remaining 65 residues were predicted by parsimonious consider-
ation of codons from the RNA alignment mapped onto the phyloge-
etic net.

Plasmid Construction and Recombinant Protein Expression

A DNA sequence encoding the predicted ancestral protein was cre-
ated based on the human chymase cDNA sequence. Four restriction
sites were introduced into the sequence for the purposes of subcloning.
The DNA sequence was constructed synthetically from 11 oligo-
nucleotides ranging in size from 82 to 76 nucleotides. Oligonucleo-
tides were annealed into three fragments with 20 nucleotide
complementary clamps. Fragments were amplified using PCR and
ligated into pGEM T-Easy (Promega). Positive clones were checked
by restriction digest and sequencing. Correct clones were assem-
bled into a psiP72 (Promega) variant with a custom polynkler. The
construct was then cloned into pIC2Za vector (Invitrogen) for ex-
pression in Pichia pastoris. Forward and reverse sequencing was
done at each cloning step. Mutants were constructed with the Quik-
Change Mutagenesis Kit (Stratagene). Recombinant proteins were
activated and purified as described previously (Chandrasekharan et al.,
1996).

Kinetic Analyses

To determine Vm (Michaelis constant) and Vmax (maximal velocity)
values for stemzyme-IDP-β and its mutants, initial velocities (v)
were determined as described by us previously (Sanker et al., 1997).
Fifteen concentrations of substrate [S] ranging between 10 and 2000
μM with enzyme were incubated at 37°C in 20 mM Tris-HCl buffer
(pH 8.0), containing 0.5 M KCl, and 0.01% Triton X-100 (final volume,
50 μl) for 20 to 60 min. For each peptide substrate, the enzyme
concentration was adjusted to ensure that ~15% of the substrate
was utilized at the lowest substrate concentration. Under these
conditions, product formation was linear with respect to time over the
duration of the incubation. Cu reverse-phase HPLC was used, as
previously described, to analyze products formed from Ang II and
its 11 analogs by stemzyme-IDP-β and its position 189 and/or 228
variants. These analyses indicated the cleavage of the X4-IS
bond in the substrate, but cleavage at other sites within the sub-
strates was not observed. Formation of the invariant product HFP
was used to calculate kinetic constants. Hydrophobic aromatic
(phenylalanine, tyrosine, and tryptophan), hydrophobic aliphatic
(leucine, methionine, valine, and alanine), basic (arginine and lysine),
acidic (glutamic acid and aspartic acid), and uncharged polar (ser-
ine) P1 residues were examined by varying the position 4 residue in
Ang II. Where individual kcat (turnover number, where kcat
values could not be determined independently due to weak binding and limited substrate solubility,
the kcat/Km ratio was determined from triplicate measurements of v
at low substrate concentration ([S] < K, where v = kcat/Km)[E][S]
(Fersht et al., 1998). [E], was determined by active site titration with
[3H]diisopropyl-phosphofluoridate.

Structural Comparisons

The S1 pocket of IDPs and other trypsin-like serine proteases were
compared by superposition of structurally conserved regions of the

following PDB files using the InsightII package (Accelrys): 3RP2 rat mast cell protease, 1KLT human chymase, 1CGH human cathepsin G, 1F18 rat granzyme B, 1FU1 human myeloblastin, 1HNE human neutrophil elastase, 1AES human cap37, 1BIT salmon trypsin, 1AHT human thrombin, 1ELV human C1S, 3TP1 bovine trypsin. The S, pocket in stemzyme-IDP-ii was modeled with Biopolymer (Accelrys) using cathepsin G as a template. Figures were constructed from transposed coordinate files using Molscript (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

Acknowledgments

We thank Robert M. Graham and David I. Martin for critical reading of the manuscript and their many helpful suggestions. This work was supported by a grant from the Australian Research Council (to A.H.). M.A.W. was supported in part by a Freedman Foundation Postdoctoral Fellowship.

Received: April 8, 2003 Revised: May 22, 2003 Accepted: June 4, 2003 Published: August 28, 2003

References


Sreenivasan, U., and Axelsen, P.H. (1992). Buried water in homolo-


Stackhouse, J., Presnell, S.R., McGeehan, G.M., Nambiar, K.P., and

Benner, S.A. (1990). The ribonuclease from an extinct bovid rumi-


maximum-likelihood method for reconstructing tree topologies. Mol.

Biol. Evol. 13, 964–969.


mony. Version 3.0. Center for Biodiversity. Illinois Natural History

Survey, Champaign, IL.


W: improving the sensitivity of progressive multiple sequence align-

ment through sequence weighting, position-specific gap penalties


tural basis for the broad substrate specificity of fiddler crab colla-


structure of the proapoptotic protease granzyme B reveals the mo-

Wyss, A. (2001). Paleontology. Digging up fresh clues about the