The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity

Sandra M. Waugh1, Jennifer L. Harris2,3, Robert Fletterick2,4 and Charles S. Craik2,4

1The Graduate Group in Biophysics, and the 2Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0466, USA. 3Present address: Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, California 92121, USA. 4Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0466, USA.

Granzyme B is a serine protease of the chymotrypsin fold that mediates cell death by cytotoxic lymphocytes. It is a processing enzyme, requiring extended peptide substrates containing an Asp residue. The determinants that allow for this substrate specificity are revealed in the three-dimensional structure of granzyme B in complex with a macromolecular inhibitor. The primary specificity for Asp occurs through a side-on interaction with Arg 226, a buried Arg side chain of granzyme B. An additional nine amino acids make contact with the substrate and define the granzyme B extended substrate specificity profile. The substrate determinants found in this structure are shared by other members of this class and help to reveal the properties that define substrate specificity.

Granzymes are a vital component of the cytotoxic lymphocyte’s ability to induce apoptosis, contributing to rapid cell death of a tumor or virally infected target cell by the cleavage of downstream substrates and the activating cleavage of caspases. Granzyme B (grB) is unique among mammalian serine proteases for its strict aspartic acid requirement in the P1 position of substrates (the amino acid N-terminal to the cleaved bond) and requires additional, extended substrate interactions (named according to Schechter) for efficient catalysis. Combinatorial substrate libraries have profiled this specificity, and indicate the preference for or absence of an amino acid at an extended position. The profile of grB is Ile or Val at P4, Glu, Met, or Gln at P3, broad preference at P2, Asp at P1, an uncharged residue at P1’ and Gly or Ala at P2’ (refs 5,6). This profile corresponds to in vitro cleavage sites, including the activation sites of caspases 3 and 7.

Ectoin (eco) is a fold specific inhibitor of chymotrypsin-like serine proteases, forming active and secondary site interactions6-10 that have been extensively characterized both enzymatically9-12 and structurally10-13. Presented here is the X-ray crystallographic structure of grB [N66Q], a variant lacking the glycosylation site, in complex with eco [81-84 IEPD], a variant reflecting the preferred grB substrate sequence at the eco active site loop P4 to P1 amino acids. This structure reveals the mechanism for recognition of Asp substrates, and the amino acids that determine extended sequence recognition.

The complex of granzyme B and ectin is a tetramer

To generate a structure of grB in a substrate-like interaction, eco [81-84 IEPD] was used to inhibit the enzyme. When the purified complex was visualized by SDS-PAGE, two distinct degradation bands appear at ~8 and 9 kDa. These fragments are the expected size if cleavage by grB occurs at the active site loop of eco [81-84 IEPD]. Although the inhibitor is cleaved, proteolytic activity measurements indicate that it still acts as an inhibitor.

The grB molecules take the three-dimensional shape of other chymotrypsin-like serine proteases (Fig. 1a), containing eight loops, labeled A-E and 1-3 (ref. 14), with the catalytic triad in the cleft between the two domains. GrB has a surface that is highly cationic, reflecting its high pi (Fig. 1b). The nucleophile, Oy of Ser 195, is 1.8 Å from the carboxyl carbon of ectin Asp 84. This sub van der Waals distance possibly represents equilibrium between a covalently bound acyl-enzyme intermediate and a substrate-like binding mode. The ectin loop C-terminal to the scissile bond is not resolved in the structure.

Each asymmetric unit contains a tetramer ~110 Å in length and 58 Å in diameter. The complex of one inhibitor and one protease is related by a two-fold nonecrystallographic symmetry axis to form the tetramer. Each ectin monomer acts as an adapter between the proteases, making two interactions, one with each protease (Fig. 2a). The primary interaction buries 1,540 Å² of grB’s surface and mimics a substrate interaction. The secondary site of interaction is ~18 Å from grB’s active site and buries 820 Å² of its surface.
The determinants of aspartic acid preference are revealed

The grB pocket N-terminal to the scissile bond, S1, holds a buried arginine in place by a network of hydrogen bonds between Gin217 and N2 and N1 nitrogen atoms of Ser221 and the N1 terminal nitrogen atom (Fig. 2b). These interactions restrain the arginine, Arg226, to the side of the binding pocket. It exposes only 12 Å2 of its solvent accessible surface when no substrate is present and creates a small patch of positive potential on the edge of S1 (Fig. 1b). Arg226 mediates substrate recognition through a salt bridge with the aspartic acid of the substrate: forming two hydrogen bonds between N1 and N2 and the carboxylate group of Asp84. Arg192 enhances the basic character of S1 and may interact with the substrate directly or through the ordered water molecule 407.

As with other members of the chymotrypsin fold, S1 is formed by loops 1 and 2. The backbone of residues 188–192, and a short, sharply bent loop between residues 221 and 224 form the walls of the pocket (Fig. 2b). The bend is formed by a cis-Pro residue and orient Arg226 towards the active site. GrB does not cleave after a Glu residue despite its preference for the Asp carboxylate18. This discrimination is due to the dimensions of the active site pocket (−7 Å in diameter and 8 Å deep). A Glu side chain is too long to be correctly oriented for catalysis and form hydrogen bonds with Arg226.

Determinants of extended specificity in grB

Extended specificity in grB comes from specific interactions with side chains of substrates in addition to the active site and backbone hydrogen bonds typical of chymotrypsin-like serine proteases (Fig. 3). S4 is a long (10 Å) and shallow (6 Å) hydrophobic groove parallel to the substrate backbone that is defined by four residues. The rings of Tyr215 and Tyr174 form the V-shaped sides, while Ile99 and Leu171 cap the ends of the pocket. Side chains longer than 6 Å will clash with the Tyr residues, giving this pocket a preference for small hydrophobic residues. The eco Ile avoids clashing with a rotomer that takes advantage of the length of the pocket. A Leu residue can be modeled into the pocket, but occurs less in combinatorial substrate assays than β-branched amino acids. Examining rotomer conformations reveals that only one rotomer of Leu makes favorable interactions with the pocket. The entropy to assume this conformation may reduce its binding efficiency. Depth and hydrophobicity are the determinants of specificity at S4 and account for the Ile and Val preference.

S3 is −6 Å long, tangential to the substrate backbone, and defined by Asn218 and the guanido group of Arg192. Asn218 forms a hydrogen bond with Glu82. Arg192 is solvent exposed and flexible, suggesting that it plays a dual role in P1 and P3 specificity. Polar residues define this pocket and a preference towards accommodating long amino acids capable of forming hydrogen bonds with one or both determinants. The experimental substrate specificity for P3 reflects this preference: Glu > Met > Gln > others.

The P2 (Pro83) side chain of ecotin points into solvent, although two grB residues make slight interactions with it. Ile99 and His57 bridge the pyrrolidone ring, each about 4 Å away. Experimentally, grB has a slight preference for Pro at this position, but accepts any amino acid except Arg or Lys. The Pro residue may orient the scissile bond for transition-state binding, while Arg residues at 192 and 226 would repel basic amino acids.

The residues of ecotin on the prime side of the scissile bond are not resolved in this structure. Determinants of the prime side specificity are suggested through modeling (Fig. 3). The catalytic triad of trypsin with wild type ecotin20 was superimposed on grB (root mean square deviation 0.126 Å). S1’ is a hydrophobic pocket adjacent to the catalytic triad defined by the disulfide bond between residues 42 and 56 and the aliphatic portion of Lys 41. All side chains are accommodated by S1’, but uncharged amino acids are experimentally preferred. This is due to the proximity of the catalytic triad and the lack of a favorable charge pair. Acidic residues at P1’ are proposed to destabilize the catalytic triad, reducing the efficiency of proteolysis. S2’ is defined by the loop D backbone and Lys 41 and Tyr 151. It narrows to 5 Å and requires an abrupt upturn of the substrate. A backbone hydrogen bond, found in ecotin111 and bovine pancreatic trypsin inhibitor (BPTI) complexes16, is formed between Lys 41 and ecotin P2’.

Specificity determinants in related proteases

GrB has an unusual specificity for Asp substrates from a sequence that is 50% identical to cathpsin G (catG), human chymase, and other granzymes. In chymase, Ala226 mediates the specificity towards large aromatic amino acids. In catG, Glu226 is held away from substrate by hydrogen bonds. Its specificity reflects the pres-
ence of the Glu residue, preferring primarily Phe, but also Lys. This suggests that the Grb active site architecture can adopt conformations specific to a range of P1 residues with the main determinant at 226 (Fig. 4). At the extended sites, Grb is very similar to the coagulation family of serine proteases. Despite their differences in overall preferred substrate (IXED for Grb and IXPR for thrombin; where X is any amino acid), their extended specificity determinants are largely the same. Residue 192 plays a key role in P3 specificity as seen by mutagenesis studies with GrbB, thrombin and activated protein C (ref. 17). Serine proteases with basic or acidic side chains at position 192 will prefer a charge complement at P3. At S4, thrombin and Grb have combinatorial substrate specificity profiles that prefer aliphatic amino acids. Hydrophobic amino acids at positions 99, 171 (180 in thrombin), 174, and 215 define the P4 pocket. These residues are easily identifiable in a sequence alignment, and indicate specificity for aliphatic amino acids at S4.

Grb [N66Q] in complex with eotin [81-84 IEFD] represents the first structure of a granzyme family member and reveals the determinants of its narrow substrate specificity. These determinants are found at 10 sites on the protease surface. Comparison with regulatory serine proteases indicates they are found throughout this subfamily of serine proteases, and suggests extended specificity can be predicted from inspection of a primary sequence. The amino acid at position 226 mediates primary specificity. At extended sites, the amino acid at position 99 determines P2 specificity, positions 192 and 218 contribute to P3 specificity, a groove formed by amino acids at positions 99, 171, 174 and 215 defines P4 specificity, and positions 151 and 41 determine P2’ specificity. These predictors depend on the high similarity of active site structure between the chymotrypsin-like serine proteases, but may provide a valuable tool to evaluate potential substrates of enzymes that have not been kinetically characterized.

Fig. 3 The extended specificity determinants of Grb. The active site of Grb [N66Q] (green) with the bound eotin [81-84 IEFD] segment (blue), shown as a, a two-dimensional (Lipplot) representation, b, a three-dimensional ribbon and c, a three-dimensional surface (InsightII, Molecular Simulations Inc.). Grb requires extended substrates for catalysis that span five recognition sites. S4 is defined by Tyr 215, Tyr 174, Ile 99 and Leu 171, and favors 3-branched hydrophobic amino acids. S3 is defined by Arg 192 and Asn 218, and favors long polar side chains. At S2, the side chain is primarily exposed to solvent but is also bracketed by Ile 99 and His 57. N-terminal to the cleaved bond, modeled substrates were used to define the specificity (light blue). S1’ is a long polar groove, and S2’ forces the substrate backbone to make an abrupt turn after a conserved hydrogen bond.

Fig. 4 A comparison of structure and sequence at the active sites of three serine proteases, grb, chymase (from 1PUP) and catG (from 1AU3). They share an active site architecture, but span a range of specificities. Grb prefers Asp substrates (yellow) with a salt bridge interaction mediated by Arg 228. Ala 228 in chymase defines a preference for large hydrophobic substrates (blue). CatG sequesters Glu 226 away from direct interaction with the substrate. It prefers Phe but also cleaves Lys substrates (pink). This figure was prepared with InsightII (Molecular Simulations Inc.).

granzyme B chymase cathepsin G
rat granzyme B 185 185 185
human chymase 203 216 224
human cathepsin G 181 181 181

nature structural biology • volume 7 number 9 • september 2000
764
Table 1 Crystallographic statistics

<table>
<thead>
<tr>
<th>Data collections statistics</th>
<th>P2₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>56.64 Å, 154.6 Å</td>
</tr>
<tr>
<td>a, b, c, β</td>
<td>57.24 Å, 119.76°</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>121,975</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>48,837</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.7</td>
</tr>
<tr>
<td>Completeness / Overall</td>
<td>90.6 % / 90 %</td>
</tr>
<tr>
<td>Rmerge (2.35–2.2 Å)</td>
<td>0.093 (0.58)</td>
</tr>
<tr>
<td>Resolution</td>
<td>42.2–2.2 Å</td>
</tr>
<tr>
<td>&lt;l&gt;/&lt;&lt;l&gt;</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Refinement statistics

<table>
<thead>
<tr>
<th>R.m.s. deviations from ideality</th>
<th>Bonds (Å)</th>
<th>0.007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angles (°)</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>Luzzati coordinate error</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>23.1</td>
<td></td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>25.6</td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>5,436</td>
<td></td>
</tr>
<tr>
<td>Water atoms</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>Average B factors (Å²)</td>
<td>42.6</td>
<td></td>
</tr>
<tr>
<td>Main chain atoms</td>
<td>43.3</td>
<td></td>
</tr>
<tr>
<td>Side chain atoms</td>
<td>47.8</td>
<td></td>
</tr>
</tbody>
</table>

<sub>Rmerge = Σ(‖Fo−Fc‖)/ΣFo</sub>
<sub>Rfree = Σ0.95(‖Fo‖−Fc) / ΣFo</sub>
<sub>Rmerge was calculated with 5% of the total reflections removed from refinement.</sub>

Methods

Protein purification. Ecoltin [81–84 IEPD], and rat grB [N66G] were prepared as described. GrB (N66G) was mixed in an equimolar ratio with ecolt in [81–84 IEPD], purified by a Mono-S cation exchange column (Pharmacia Biotech, Uppsala Sweden) from 50 mM to 1.5 M NaCl, and concentrated to > 5 mg ml<sup>-1</sup>.

Crystallization. Crystals, averaging 50 µm x 100 µm x 500 µm, were grown at 18 °C by hanging drop vapor diffusion. Drops (3 µl: 1: 1 ratio of protein and precipitating solution) were equilibrated against 1 ml of precipitating solution (100 mM Na acetate pH 5.8, 100 mM ammonium sulfate, and 25% (w/v) PEGmme 2000.) Crystals were frozen for data collection with no alterations to the solution.

Data collection, processing and molecular replacement. Two data sets were measured under cryogenic conditions at the Stanford Synchrotron Radiation Laboratory on beamline BL7-1 with a MAR imaging plate. Crystals belonged to the space group P2₁. Data sets were processed with MosFlm<sup>16</sup> and scaled and assigned the same free R using CCP4<sup>15</sup>. A program for crystallographic molecular replacement by evolutionary search (EMPR-Aagoun Pharmaceuticals, La Jolla, California) sequentially fit an ecoltin dimer (from 1AZZ) and two cathepsin G molecules (from 1AUG), modified to Ala where they differed from grB, into the unit cell.

Refinement. Initial rigid body rotation and minimization were performed to 2.5 Å with CNS1.0 (ref. 21). A 2.5 Å F<sub>2</sub>− F<sub>1</sub> map was used to rebuild all side chains. Strict noncrystallographic symmetry (NCS) constraints were used until the R-factor reached ~28% (R<sub>free</sub> ~31%). The NCS constraints were removed and cycles of rebuilding and refinement using CNS1.0 and quanta98 (Molecular Simulations Inc., San Diego, California) were alternated with rounds of simulated annealing and addition of solvent molecules. The final model was verified using PROCHECK<sup>21</sup>, with 87% of residues in the most favorable regions of the Ramachandran plot.

Coordinates. Coordinates of ecolt [81–84 IEPD] in complex with grB [N66G] have been deposited in the Protein Data Bank (access code 1F80).

Acknowledgments

This work was supported by grants to C.S.C. from the National Science Foundation and National Institutes of Health, to R.J.F. from National Institutes of Health, and to J.H.L. from the National Institutes of Health Biotechnology Training Grant Fellowship.

Correspondence should be addressed to C.S.C. email: crik@cgl.ucsf.edu or S.M.W. email: wasugh@mutant.ucsf.edu

Received 9 May, 2000; accepted 28 July, 2000.