nature
structural biology

The oligomeric structure of human granzyme A is a determinant of its extended substrate specificity

Jessica K Bell[1,4], David H Goetz[1,4], Sami Mahrus[1,2], Jennifer L Harris[1,2], Robert J Fletterick[3] & Charles S Craik[1]

The cell death–inducing serine protease granzyme A (GzmA) has a unique disulfide-linked quaternary structure. The structure of human GzmA bound to a tripeptide CMK inhibitor, determined at a resolution of 2.4 Å, reveals that the oligomeric state contributes to substrate selection by limiting access to the active site for potential macromolecular substrates and inhibitors. Unlike other serine proteases, tetrapeptide substrate preferences do not correlate well with natural substrate cleavage sequences. This suggests that the context of the cleavage sequence within a macromolecular substrate imposes another level of selection not observed with the peptide substrates. Modeling of inhibitors bound to the GzmA active site shows that the dimer also contributes to substrate specificity in a unique manner by extending the active-site cleft. The crystal structure, along with substrate library profiling and mutagenesis, has allowed us to identify and rationally manipulate key components involved in GzmA substrate specificity.

Granzymes (Gzms) are serine proteases secreted from cytotoxic T lymphocyte (CTL) and natural killer (NK) cell granules in response to signals on virally-infected and tumor cells. Upon uptake by the target cell, Gzm mediated proteolytic cleavage of cellular substrates leads to rapid target cell death[1]. Among the 13 known members of Gzms (A–M)[2,3], GzmA is the most abundantly expressed in CTLs[4]. Unlike GzmB, GzmA works in a caspase-independent manner, providing a failsafe mechanism for apoptosis when viral or tumor cell components are able to neutralize the caspase pathway[4,5]. A hallmark of GzmA-mediated cell death is single-stranded DNA nicks through the activation of NM23-H1, a nucleoside diphosphate kinase implicated in suppression of tumor metastasis[6,7]. NM23-H1 is released when GzmA cleaves NM23-H1's inhibitor, nucleosome assembly protein SET[6,7]. Other identified substrates include pro–IL-1β[8], thrombin receptor[9], lamin b[10], histone H1, histone H2b[11], SET-associated DNA bending protein HMG2[12], and ApeI[13], a multifunctional protein involved in base excision repair.

GzmA is a member of the S1-A (trypsin-fold) family of serine peptidases. The primary specificity pocket (S1) favors basic residues Arg and Lys[14]. Unlike GzmB, where a tripeptide substrate is required to observe cleavage and is maximal with an idealized tetrapeptide, GzmA readily cleaves single amino acid substrates[2]. Although recognition of extended substrate sites is not necessary for cleavage, the identification of specific substrates for GzmA indicates that its cleavage occurs discriminately. The ability of GzmA to target specific substrates is further supported by comparing trypsin and GzmA digests of nuclear lysates, which showed dramatic differences in specificity[15]. Trypsin degraded the lysate within 1 hour whereas GzmA degraded only 1 of 22 bands within 25 hours[15]. GzmA's substrate selection may also be influenced by the enzyme's unique ability to form a disulfide-linked homodimer[16]. Using the standard chymotrypsinogen amino acid numbering system[17] for GzmA, the unique reactive cysteine at position 93 is in addition to the eight cysteines forming the four disulfide bonds normally found in the trypsin-fold proteases. The proximity of the resulting dimer interface to the substrate-binding pocket could impose further constraints on macromolecular substrate selection.

Comparative analysis of structurally similar but functionally different enzymes such as trypsin and chymotrypsin has traditionally been used to identify key determinants of substrate specificity at the S1 site. With the recent development of methods to explore the extended specificity pockets of proteases, the basis for molecular recognition can be analyzed beyond the S1 site. In this respect, another key to deciphering GzmA's specificity may lie in small differences between it and closely related proteases. The mouse GzmA shares 70% identity with its human ortholog. However, experimentally the differences between mouse and human GzmA extended substrate specificity had not been previously studied.

To identify the components responsible for GzmA substrate selection, we have structurally characterized the dimeric human GzmA bound to a tripeptide CMK inhibitor. We have further probed substrate selection by GzmA, using positional scanning synthetic combinatorial libraries (PS-SCL)[18–20] to map the extended substrate specificity at the S2–S4 pockets of human GzmA and its mouse homolog. Interestingly, although these orthologs share the same P1 specificity, the extended specificity differs. Given the combination of sequence similarity and extended substrate specificity differences, we have used our atomic structure and mapped potential substrate specificity switches to residues lining the extended substrate binding

[1]Department of Pharmaceutical Chemistry. [2]Graduate Program in Chemistry and Chemical Biology. [3]Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA. [4]These authors have contributed equally to this work. Correspondence should be addressed to C.S.C. (craik@cgl.ucsf.edu).

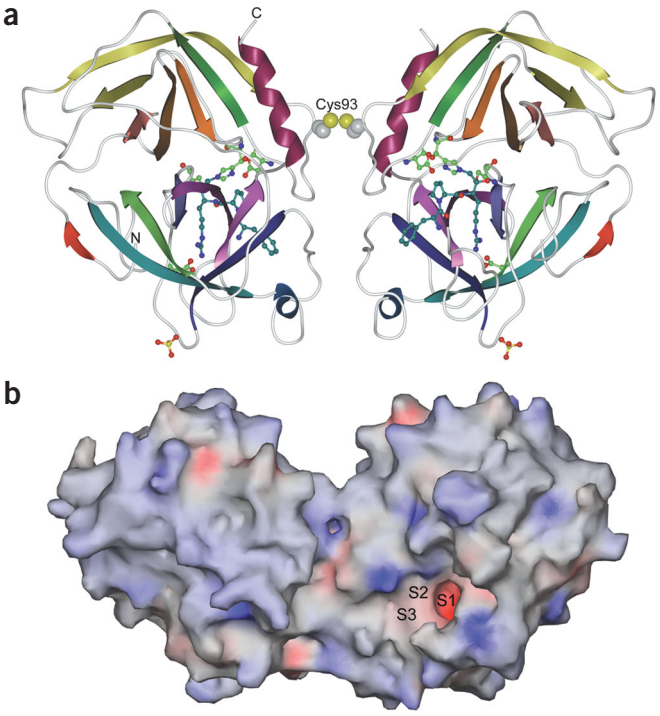

Figure 1 Overall structure of dimeric human granzyme A. (**a**) Ribbon diagram of dimeric human GzmA. The secondary structural elements are colored in a gradient from N to C terminus. The Ser195-His57-Asp102 catalytic triad and P1 coordinating Asp189 are shown in ball and stick (carbons, light green) as is the D-Phe-Pro-Arg-chloromethylketone (carbons, cyan). The disulfide linkage is depicted as space-filling model for residue 93 and its symmetry mate. A sulfate ion contributed from the crystallization solution is bound at the base of loop 184-B–197 by Arg186 and Arg188. (**b**) Surface representation mapped with potentials shows the overall positive charge of the molecule reflected by its pI > 9 and the distinct negative charge of Asp189 emanating from the S1 pocket.

pockets and confirmed our predictions through mutagenesis and PS-SCL characterization.

RESULTS

Overall structure

The structure of human GzmA bound to an irreversible D-Phe-Pro-Arg CMK inhibitor was determined by molecular replacement (**Fig. 1a,b**). The asymmetric unit contains a single subunit with the disulfide linkage of the homodimer lying on a crystallographic two-fold symmetry axis. Each GzmA monomer contains the classic trypsin-fold with two β-barrel domains, each consisting of six β-strands. The two domains are connected by a long flexible loop between residues 109–134 with the catalytic triad of Ser195, His57 and Asp102 nestled into the domain crevice (**Fig. 1a**). Although GzmA is most similar to GzmK (45% sequence identity) and an activated mutant of complement factor D[21] (42% identity), it is structurally most related to trypsin (*Z*-score 47.9), followed by GzmB (*Z*-score 34.4) and chymase (*Z*-score 34.3) as assessed by the Dali server[22]. The hierarchy of structural similarity is primarily because the active sites of both complement factor D and GzmK were structurally characterized in their inactive conformations.

Architecture of the dimer interface

The dimer interface, aside from the disulfide formed between symmetry-related Cys93 residues, is formed by hydrophobic interactions and hydrogen bonds between symmetry-related loops (**Fig. 2**). In this view, the top of the interface is capped by the solvent-exposed disulfide-linkage between the symmetry-related Cys93. A water bridge forms between Wat94 and the carbonyl oxygen of Cys93 and Nε amino group of Lys233-B′ (The prime indicates the residue is from the second subunit in the dimer, while the A, B, C or D notation indicates residue numbering insertion code to maintain the trypsin numbering scheme throughout the structure). The carboxylate of Asp95 can also make a hydrogen bond with δN of Asn179 through Wat29. The side chain of Met178-A occupies two conformations at the interface. One conformation packs against the Cγ of Arg165, Cγ2 of Val175-C and Cβ of Phe174′. The alternate conformation protrudes prominently into the dimer interface making hydrophobic contacts with Cγ2 of Thr98′ and Cβ of Val175-C′. Val175-C also interacts with its symmetry mate across the crystallographic two-fold axis. The side chain of Lys166 interacts through its aliphatic carbons with Phe174′ and via its Nε amino group to the carbonyl oxygen of Asn173′ to form the base of the dimer interface. Phe174 also interacts with the aliphatic side chain of Arg165′.

The total buried accessible surface area between the monomers at the dimer interface is 1,140Å^2 or 5% of the total surface area of each monomer. The magnitude of the buried surface area, the degree of charge complementarity, and the number of direct and water-mediated hydrogen bond partners across the dimer interface argue that in the absence of the disulfide bond, the protein may exist either primarily as a monomer or in a monomer-to-dimer equilibrium in solution. The dimer interface observed in the crystal structure is, however, not likely to be an artifact of crystal packing conditions because interactions in the area bound by Met178-A, Lys166, and Lys166′, although covering limited surface area, are extensive. To address this point, Cys93 was replaced with a serine and the variant was analyzed by size-exclusion chromatography. Gel-filtration analysis of the C93S GzmA demonstrated that without the disulfide bond this variant eluted as a monomer clearly differentiated from the dimer peak of the wild-type enzyme (data not shown). Therefore, the disulfide bond appears necessary to form a stable dimeric enzyme. *In vivo* the oligomeric state may be further stabilized by the association of serglycans to the enzyme surface protecting both the solvent-exposed disulfide from a reducing environment[1,23] and limiting access of macromolecular inhibitors, such as anti-thrombin III[24].

Subsites C-terminal to the scissile bond

The CMK inhibitor bound to GzmA only defined the subsites N-terminal to the scissile bond. The substrate binding pockets C-terminal to the scissile bond were evaluated based upon the structural comparison of GzmB bound to ecotin, a macromolecular trypsin-fold serine protease inhibitor (PDB entry 1FI8), to the GzmA structure. The close structural relationship of GzmA to GzmB allows these substrate pockets to be inferred using ecotin's extended β-strand bound in GzmB's active site superimposed onto the GzmA active site. The substrate binding pockets of proteases, termed subsites (S), are defined by their position relative to the scissile bond. The subsites that recognize and bind to residues in the substrate C-terminal to the scissile bond are designated S′ to differentiate them from the subsites N-terminal to the scissile bond designated as S. The subsites are numbered starting with 1 for the first amino acid before/after the scissile bond. The corresponding amino acids of the substrate (P) that fit into these pockets have the same nomenclature.

From our analysis, loops Leu35–Cys42 and His57–Ser63 would form the S1′ pocket (**Fig. 3a**). This pocket exists as a shallow hydrophobic groove. The S1′ pocket appears well suited for binding

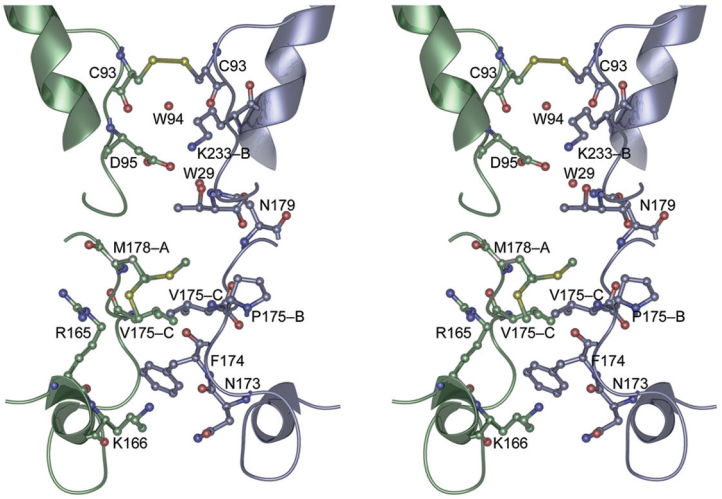

Figure 2 Dimeric interface of human granzyme A. A close-up view of the dimer interface in the same orientation as **Figure 1**. The ribbons and carbon atoms of the two subunits are colored green and purple, respectively. Only unique subunit–subunit interactions are shown, symmetry equivalent interactions were removed for clarity. Met178-A residues are in two conformations (0.5 occupancy each).

substrates with long linear hydrophobic amino acids such as the aliphatic side chain of lysine or methionine at P1′. Alternatively, unbranched hydrophobic (Ala) or polar residues (Ser) can be accommodated. The S2′ pocket is formed by a cleft between Gly142–Trp151 and Leu35–Cys42 (**Fig. 3a**). The pocket formed by these segments is larger than the S1′ pocket yet it is also hydrophobic and appears well suited for binding to branched hydrophobic or aromatic residues. No specificity pockets beyond S2′ appear defined. In addition, to avoid steric clashes with residues Ser37–Ile40 and Trp151, substrate must turn sharply out of the active site cleft after P2′ (**Fig. 3a**). The requirement of a sharp turn may preclude amino acids with limited conformational flexibility, such as proline, at P1′ or P2′.

The PS-SCL technology for experimentally defining sequence preferences in protease subsites is currently limited to positions N-terminal to the scissile bond, although synthetic combinatorial methodology for routine mapping of the sites C-terminal to the scissile bond is under development[25]. Since direct experimental data could not readily be obtained, we have compared our predicted prime side preferences, P1′ = K/M or A/S and P2′ = V/L/I/aromatic, with known cleavage sequences (**Table 1**). The P1′ predicted and known cleavage sequences are consistent with a slight preference for serine in the macromolecular substrate sequences. The P2′ position in known cleavage sequences is more varied than expected from our predicted preference of primarily hydrophobic residues.

Subsites N-terminal to the scissile bond

The N-terminal subsites of GzmA were defined structurally by the chloromethylketone inhibitor, D-Phe-Pro-Arg-CMK, irreversibly bound in the GzmA active site and experimentally through PS-SCL. With PS-SCL, an enzyme is screened against an arrayed oligopeptide library coupled to a reporter group, in our case the highly sensitive 7-amino-4-carbamoyl methyl coumarin (ACC) fluorescent group. These libraries provide an unbiased screen to identify preferred peptide substrates for a given protease. Determination of the preferred residues is accomplished readily using a library in which the P1–P4 positions are systematically arrayed as one of the amino acids commonly found in proteins excluding cysteine and methionine and including norleucine as a methionine mimic. By combining both the structural and experimental data, we were able to develop an understanding of sequences that can be accommodated within the GzmA subsites N-terminal to the scissile bond.

In the crystal structure, residues Asp189–Ser195, Gly224–Val227, Lys219-B–Gly221-A, and Ser214–Leu217 form the S1 pocket (**Fig. 3a**). The S1 pocket is the most well-defined substrate interaction site in the active-site cleft (**Fig. 3b**). The specificity for a basic residue of the S1 pocket is primarily due to the side chain of Asp189 at the bottom of the pocket, which is positioned for favorable short-range electrostatic interaction with an arginine or lysine as the P1 residue. The guanidinium group of the CMK inhibitor's arginine, bound in the S1 pocket, forms two direct hydrogen bonds with Asp189. Similar to trypsin, a solvent molecule, Wat63, in the S1 pocket makes hydrogen-bonds to the Nε atom of the P1 arginine side chain. The irreversible inhibitor is covalently linked to the Oγ of the Ser195 and also stabilized through the Nε of His57 to the hydroxyl of the former ketone group (**Fig. 3c**).

Activity against substrates in a PS-SCL in which the P1 residue of a tetrapeptide sequence was systematically fixed as one of the 19 (excluding methionine and cysteine and including norleucine) proteinogenic amino acids (P1-diverse) confirmed that GzmA possessed primary specificity for basic amino acids, with arginine preferred over lysine (**Fig. 4a,b**). The human and mouse GzmA homologs were both evaluated by PS-SCL as a means to verify GzmA extended substrate specificity. The isozymes were expected to give near-identical substrate preferences due to their high sequence similarity (~70%).

The S2 pocket of GzmA is formed by the Asp95–Arg99 and Ser214–Leu217 loops (**Fig. 3a**). The CMK inhibitor co-crystallized with GzmA has a proline at the P2 position. The proline carbonyl makes hydrogen bonds to a solvent molecule within the active site while the side chain forms hydrophobic interactions with the aliphatic

Table 1 Human granzyme A PS-SCL substrate sequence compared to *in vivo* substrate cleavage sequences[a]

		P4	P3	P2	P1	↓	P1′	P2′			
PS-SCL		I/V	G/A/S	N/D	R						This study
pro–IL-1β	D	A	P	V	R	↓	S	L	N	C	T	Ref. 8
Thrombin receptor	T	L	D	P	R	↓	S	F	L	L	R	Ref. 9
Histone H1	K	L	G	L	K	↓	S	L	V	S	K	Ref. 11
Histone H2b	A	P	A	P	K	↓	K	G	S	K	K	Ref. 11
SET	Q	T	Q	N	K	↓	A	S	R	K	R	Ref. 6
Lamin B	V	T	V	S	R	↓	A	S	S	S	R	Ref. 10
HMG2	E	D	M	A	K	↓	S	D	K	A	R	Ref. 12
Apel	K	T	A	A	K	↓	K	N	D	K		Ref. 13

[a]The ↓ indicates the scissile bond.

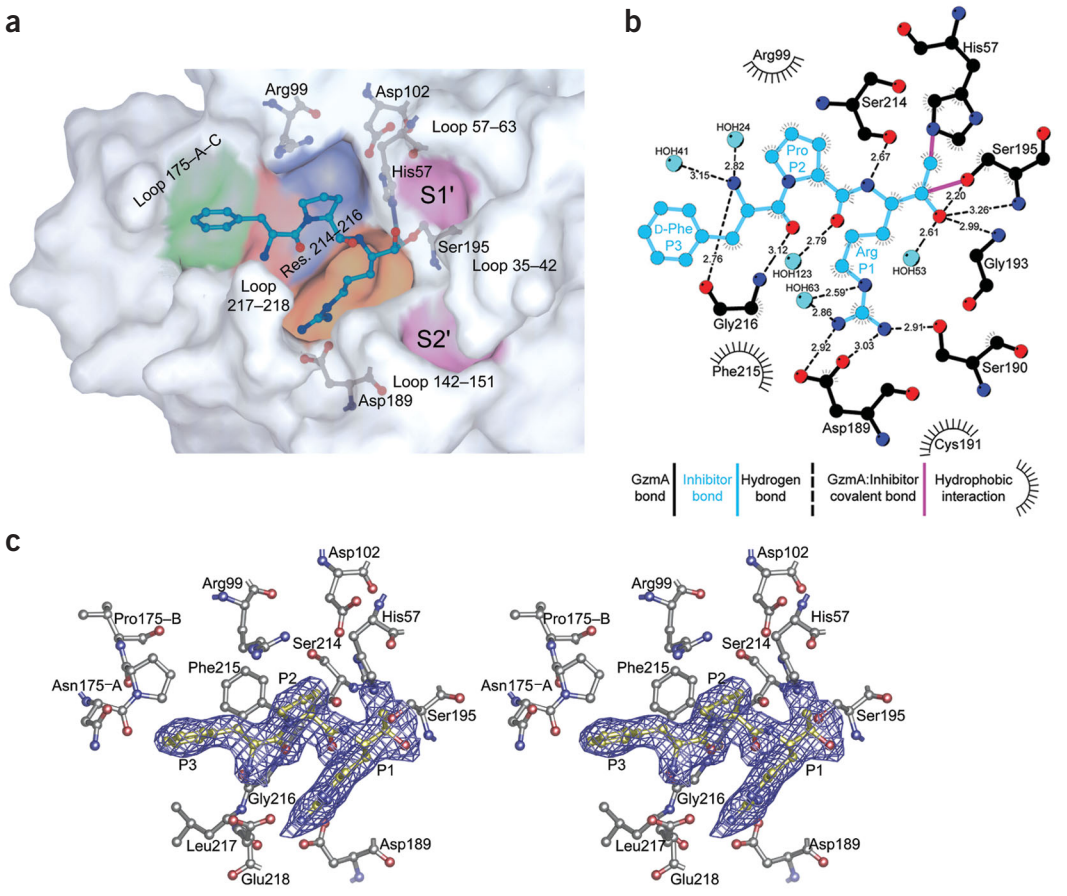

Figure 3 The active site of human granzyme A. (**a**) Ball-and-stick representation of the bound inhibitor, D-Phe-Pro-Arg-CMK (carbons, cyan) and residues that frame the substrate binding pocket depicted in the context of the molecular surface. The molecular surfaces of the proposed S1′ and S2′ subsites are colored in magenta; S1 subsite, orange; S2, blue; S3, red; S4, green. (**b**) Ligplot representation showing direct interactions between GzmA and the bound inhibitor. D-Phe-Pro-Arg-CMK bonds and carbons are cyan. Bonds between the irreversible inhibitor and GzmA are magenta. (**c**) Stereo view of the refined ($2F_o - F_c$) electron density for the CMK inhibitor (carbons, yellow) bound to the GzmA active site (carbons, gray).

side chain of Arg99 (**Fig. 3b**). The guanidinium group of Arg99 caps the S2 pocket favoring small amino acids capable of hydrogen bonding with the basic side chain.

Previous experimental data for P2 subsite sequences reported a preference for phenylalanine at this position[24,26–28]. The P2 mapping by PS-SCL, as well as subsequent P3 and P4 mapping, utilized a tetrapeptide library in which the P1 position was held constant as an arginine (P1-Arg) and the P2 position (or P3 or P4 position in subsequent mapping) varied systematically. Our observed PS-SCL P2 specificities (Asn for human, Phe for mouse, **Fig. 4a,b**) are consistent with the literature. Asparagine does not appear to have been previously tested at the P2 position. The PS-SCL results for the P2 pocket were our first indication that the small sequence variations between human and mouse homolog might lead to differences in tetrapeptide substrate selection. Our structure shows that Arg99 is the primary determinant for P2 specificity, contrary to previous speculation based on models that the role of Arg99 was of minimal importance[29]. Changes in the loop containing this residue could lead to subtle differences in Arg99 side chain positioning. To account for variation in residue preference between the mouse and human GzmAs at P2, we modeled the mouse active-site cleft, based upon the human structure (data not shown). The model suggests that the Pro96 to Glu96 substitution (human to mouse) leads to increased flexibility of the Arg99 loop. For the mouse homolog preference of phenylalanine, the guanidinium group of Arg99 may be positioned to interact favorably with the π-electrons of phenylalanine.

The S3 pocket of GzmA is much narrower in comparison to the digestive serine protease trypsin. In this region, a two-residue insertion, 218 and 219-A, results in residues 217–220 looping up from the floor of the S3 pocket (**Fig. 3a**). This loop forms a wall that greatly constricts the length of any P3 side chain. The S3 pocket is thus suited for binding small amino acids.

Our experimental evaluation of human GzmA tetrapeptide specificity is consistent with our structural prediction and previous reports[28] (**Fig. 4a**). The mouse homolog P3 preference is strikingly different from the human enzyme (**Fig. 4b**). As in many serine protease structures, the P3 side chain can form an interaction with the base of the 218–220 loop and solvent. The mouse enzyme's preference for aromatic amino acids may be explained by increased flexibility in the 218–220 loop due to a Leu→Gly substitution at residue 217 and a decrease in the size of the loop due to a deletion at residue 219-A. This change has the effect of preserving the hydrophobicity of the P3 pocket while greatly increasing its size.

Residues His171–Gly177, Pro225–Val227, and Phe215–Leu217 form the S4 pocket (**Fig. 3a**). These residues form a shallow hydrophobic groove, similar to trypsin that is best suited for small hydrophobic amino acids. The CMK inhibitor bound in the active-site cleft has a D-Phe at P3. The D-amino acid has the effect of positioning the side chain of P3 into the S4 pocket leaving the S3 pocket unoccupied except for the backbone of the inhibitor which forms hydrogen bonds to both the GzmA mainchain and solvent molecules (**Fig. 3b**). The D-Phe is too bulky to nestle down into the hydrophobic S4 pocket further supporting our hypothesis that GzmA prefers small hydrophobic amino acids at S4.

The P4 subsite PS-SCL results for both human and mouse indicate an overall preference for hydrophobic residues consistent with the

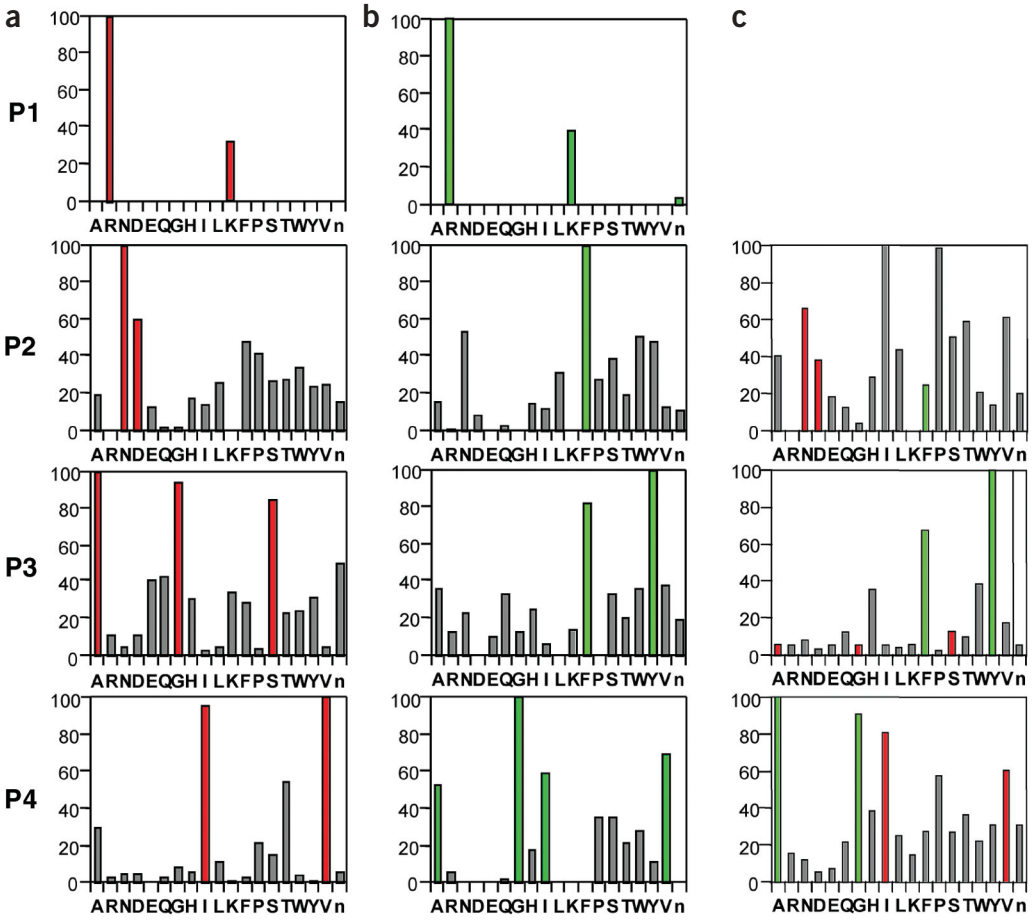

Figure 4 Combinatorial substrate library results for human, mouse and H→M granzyme A. A set of three histograms for each variant represents the P1, P2, P3 and P4 positions within the substrate libraries. For each graph the *y*-axis represents the normalized rate of substrate cleavage (fluorophore release) over time and the *x*-axis represents the amino acid positioned at P1, P2, P3 or P4. The positions in the substrate that are not held constant contain an equimolar mixture of 19 amino acids (Cys and Met excluded, Nle included). (**a**) Human GzmA. (**b**) Mouse GzmA. (**c**) H→M GzmA. The red and green colored bars represent preferred amino acids at a given subsite for human and mouse GzmAs, respectively.

structural analysis (**Fig. 4a,b**). The mouse enzyme's P4 profile is unusual in that glycine is preferred, but valine and isoleucine are also tolerated. This appears to be due to the two substitutions Gly209→Trp and Ile201→Gly (from human to mouse). Both the isoleucine side chain in human and the tryptophan side chain in mouse GzmA occupy equivalent positions within the floor of the S4 site. The larger tryptophan side chain leaves slightly less room for a bulkier P4 amino acid, such as the human homolog's preferred valine or isoleucine versus the mouse preferred glycine.

From the PS-SCL data of the human and mouse GzmA homologs we identified variations in extended substrate sequence preferences. We confirmed the absolute preference by the kinetic parameters for two substrates, designed based on the profiles from the PS-SCL. The preferred human GzmA substrate, Ac-VANR-ACC, showed a ~130-fold selective activity for human GzmA over mouse GzmA, with k_{cat} / K_m values of 34,000 ± 7,000 M^{-1} s^{-1} and 260 ± 30 M^{-1} s^{-1}, respectively. The preferred mouse GzmA substrate, Ac-GYFR-ACC, showed a four-fold selective activity for mouse GzmA over human GzmA with k_{cat} / K_m values of 36,800 M^{-1} s^{-1} and 9,500 ± 900 M^{-1} s^{-1}, respectively.

Combining our analysis of the crystal structure, differences between the PS-SCL for the human and mouse variants and small sequence variations between the two homologs that could account for unique substrate preferences, we designed a human-to-mouse mutant (H→M) of the human homolog that should have conferred mouse P3, P4 substrate preferences. Within the S3 pocket, Leu217 was deleted and residues 218 and 219-A were mutated to match the mouse sequence (Glu218→Gly, Asn219-A→Glu). The size of the S4 pocket was reduced by mutating Gly209 to a tryptophan. The results of the PS-SCL experiments showed that indeed the P3 preference has been completely converted to phenylalanine and tyrosine versus Gly/Ala/Ser (**Fig. 4c**). The P4 selection has now broadened to include Ala/Gly/Ile versus the human homolog's strict preference for Val/Ile. Interestingly, the P2 profile seems to have lost its preference for either the human or mouse predicted sequence. One explanation may be that the hydrogen bonds between the Ser214 and Gly216 to the substrate backbone have been disrupted or reordered due to mutations of nearby or neighboring residues (217, 218, 219-A). Despite the broadened substrate sequence at P2, we have shown that specificity can be manipulated by insertion/deletions in loops forming substrate-binding pockets or by side chain substitutions to alter pocket volume and polarity. The results demonstrate that significant, directed alteration of the specificity of GzmA can be achieved with just four changes in residues Leu217, Glu218, Asn219-A and Gly226.

Dimerization and substrate specificity

To probe whether dimerization is a determinant of extended substrate specificity, variants of mouse and human GzmA that lacked the intermolecular cysteine amino acid, C93S, were screened by PS-SCL. The resulting profiles of the cysteine variants were identical to those of the wild-type enzymes (data not shown), demonstrating that dimerization does not directly determine the P1–P4 substrate specificity for peptide substrates. Variants lacking the intermolecular disulfide bond also have kinetic parameters equivalent to those of their wild-type counterparts (data not shown).

Several macromolecular substrates for GzmA have been identified from a combination of *in vivo* and *in vitro* experiments[6–13]. We com-

pared the results of the PS-SCL experiments for human GzmA to the published cleavage sequences of macromolecular substrates (**Table 1**). In contrast to enzymes like GzmB[30,31] and the caspases (ref. 32 and references therein), whose synthetic substrate profiles match with identified *in vivo* cleavage sites, little agreement outside the P1 site can be seen between the PS-SCL results and the native substrates of GzmA. The variation between peptide and macromolecular cleavage sequences strongly supports our hypothesis that the dimerization of GzmA dominates the substrate selection through steric discrimination with macromolecular substrates in a manner analogous to tryptase[33]. In this context the molecular determinants for P2–P4 specificity, which we have identified and are readily manipulated by mutagenesis, may have little effect on *in vivo* macromolecular substrate selectivity. Indeed, although the PS-SCL data on mouse and human GzmA are very different, the cleavage sequences within identified natural substrates have clearly not co-evolved to satisfy the optimal P2–P4 specificity of GzmA in each species. Obviously, the interactions outside of the S1–S4 pockets affect substrate selection suggesting a reliance upon access to the active site cleft. We conclude that although GzmA evolved to bind a preferential sequence within its S1–S4 pockets, the context of that sequence within a macromolecular substrate must also complement the oligomeric state of GzmA.

Modeled interaction of GzmA with macromolecular inhibitors

The molecular surface of the GzmA dimer reveals an active site partially sequestered by the proximity of the dimer interface. During crystallization trials of GzmA, ecotin was tested as a potential inhibitor and macromolecular marker of the substrate binding sites, as had been previously described[31,34]. Neither the native nor a monomeric variant[35] of ecotin have shown any effect on GzmA activity up to a 1,000× excess of inhibitor. Classically, ecotin forms extensive contacts both in the active site and through a secondary binding site[36]. To confirm our hypothesis that the dimeric structure of GzmA resulted in limited access to the active site, we superpositioned the rat GzmB–ecotin coordinates (PDB entry 1FI8) to a monomer of GzmA, using a GzmB subunit (r.m.s. deviation between GzmB and GzmA is 1.09 Å over 214 Cα atoms). In the model of the resulting GzmA–ecotin complex, ecotin sterically clashes with both subunits of GzmA. The steric clash arises from the subunit of the dimeric ecotin not directly bound to the active site in the superposition. This ecotin subunit completely overlaps with GzmA subunit A. The ecotin subunit modeled into the GzmA active site overlaps the Leu35–Cys42 loop involved in the forming both the S1′ and S2′ pockets (refer to **Fig. 3a** for the position of the Leu35–Cys42 loop), explaining why a monomeric variant of ecotin is also unable to inhibit GzmA. The clash arises primarily from the disulfide bond between ecotin residues 50 and 87. This disulfide bond has the effect of stabilizing ecotin at the P2′ position in a conformation that appears to be incompatible with the aforementioned requirement of a sharp bend out of the GzmA active-site cleft C-terminal to the scissile bond.

The Kunitz type serine protease inhibitor bovine pancreatic trypsin inhibitor (BPTI) was also tested for its ability to inhibit GzmA and shown to be ineffective. The model of the GzmA–BPTI complex obtained by superimposing trypsin of the trypsin–BPTI complex (PDB entry 1BRB) onto a GzmA subunit (r.m.s. deviation between trypsin and GzmA is 1.2Å over 215 Cα atoms), revealed a collision with the Leu35–Cys42 loop in the S2′ pocket, similar to the GzmA–ecotin model. Additionally, in the GzmA–BPTI model, residues 10–12 of BPTI overlap with the 215–218 loop that forms the wall of the S3 pocket (refer to **Fig. 3** for the position of the 215–218 loop). The overlap is due to the unusually tight specificity of GzmA for small amino acids at P3.

Tsuzuki *et al.*[37] recently reported that GzmA binds to the Kazal type inhibitor pancreatic secretory trypsin inhibitor (PSTI) with a K_i of 34 ± 7 nM. A model of GzmA–PSTI interaction showed clashes in the rigidly modeled complex, but they are slight compared to the BPTI–GzmA and ecotin–GzmA models. Interestingly, residues 12–14 of the modeled PSTI in the active-site cleft of GzmA occupy a unique position by looping around the S4 pocket and occupying a possible S5 cleft between the Asp95–Arg99 and Phe174–Val175-C loops (refer to **Fig. 3a** for the position of these loops). The P5 residue of PSTI (Glu12) in the model shows that this residue could also interact with the adjacent dimer interface, consistent with our hypothesis that macromolecular substrates and, in this case, inhibitors can utilize the extended surface of the dimer to make favorable interactions.

Structural comparison

Comparison of GzmA to its closest sequence homologs reveals the residues responsible for making GzmK and complement factor D inactive. The crystal structure of pro-GzmK (PDB entry 1MZA) differs from that of human GzmA most strikingly in the loop between Ser214 and Cys220. In GzmK residue 215 is a glycine instead of the aromatic residue in GzmA. A glycine at position 215 disrupts the canonical β-structure of this loop, shifts the main chain into the active-site cleft and precludes binding to substrate in pro-GzmK.

Complement factor D is thought to maintain a catalytically inactive structure until bound to its natural substrate C3bB. The critical residues responsible for the inactive conformation of factor D appear to be Ser215 (Phe215 in GzmA) and Ser217 (Leu217 in GzmA). In GzmA the aromatic side chain at position 215 interacts with the side chains of residues Arg99, Tyr172, and Leu227. At position 217, the leucine side chain primarily interacts with Tyr172. These hydrophobic packing interactions have the effect of stabilizing the loop into a canonical β-sheet thus forming the P1–P3 active site cleft. The hydrophobic to serine substitutions in the factor D sequence result in increased flexibility of this loop, precluding productive packing interactions with the side chains of neighboring residues. As a result, His57 of factor D adopts the non-productive *gauche* conformation to avoid a steric clash with Ser215. Ser94 of factor D also contributes to the inactive conformation of His57 by providing room for the *gauche* conformation. In GzmA residue 94 is a tyrosine that would sterically limit the alternate conformation of His57.

DISCUSSION

The unique quaternary arrangement of dimeric GzmA recycles the classic serine protease trypsin-fold while permitting a unique mechanism of substrate specificity by (i) steric hindrance between the second subunit of GzmA and certain macromolecular substrates, and (ii) providing a continuous extended active site cleft across the dimer interface. The two roles proposed for the dimeric enzyme may work synergistically to exclude non-specific or detrimental interactions via steric hindrance while enhancing the S1–S4 binding pocket by extending the potential interaction surface.

The structure of homodimeric GzmA has also recently been determined by Hink-Schauer *et al.*[38] The overall quaternary arrangement of the structure is identical to the structure we report, despite different crystallization conditions and unit-cell packing. The equivalent dimer interface in the two reported structures further supports our conclusion that the observed quaternary arrangement is the physiological dimer. The report by Hink-Schauer *et al.*[38] also proposed, in agreement with our conclusion, that this unique quaternary arrangement provides an additional level of substrate specificity that supercedes that of the active-site cleft. However, Hink-Schauer *et al.* state that the

Table 2 Data collection and refinement statistics

Data collection	
Resolution (Å)	2.4
Reflections	
Total	175,040
Unique	13,373
Completeness[a]	0.99 (0.94)
$I / \sigma<I>$[a]	14.3 (2.3)
R_{merge} (%)[a,b]	10.3 (41.7)
Refinement statistics	
R_{cryst} (%)[c]	19.1
R_{free} (%)[c]	23.2
Number of atoms	
Protein	1,778
Ligand	30
Water	129
Average *B*-factor (Å^2)	
Protein	35.8
Ligand	29.5
Water	39.7
R.m.s. deviation	
Bond lengths (Å)	0.013
Angles (°)	1.544
Ramachandran (%)	
Most favored	82.2
Allowed	17.8

[a]Values for the highest resolution shell (scaling and reduction = 2.49–2.4 Å, refinement = 2.46–2.4 Å) are given in parentheses. [b]$R_{merge} = (\Sigma |I - <I>|) / (\Sigma I)$ where I is the measured intensity of reflections with indices *hkl*. [c]$R_{cryst} = (\Sigma||F_o| - |F_c||) / (\Sigma |F_o|) \times 100$ where F_o and F_c refer to the observed and calculated structure factor amplitudes for indices *hkl*, respectively: R_{free} set contained 5% of the total data.

active-site cleft of GzmA exhibits no specificity beyond the S1 pocket[38], in contrast to our PS-SCL results. The reason for this inconsistency is likely due to their reliance on substrate specificity data derived from a limited number of thiobenzyl ester substrates[2] and phosphonate inhibitors[39]. This discrepancy highlights another advantage of the PS-SCL technology used in our assays.

The solvent exposed disulfide-linkage may also provide a mechanism for a specificity and/or activity switch for GzmA that is unique among the trypsin-fold serine proteases. In an oxidizing environment such as storage granules and the extra-cellular milieu, GzmA would have a macromolecular specificity due to dimerization that overrides that of the monomer, perhaps protecting the CTL or extracellular environment from renegade cell death–inducing effects of GzmA. In the reducing environment of the target cell cytosol, however, GzmA may exist as a monomer allowing greater access to macromolecular substrates. *In vivo* studies will shed more light on the role of GzmA's unique oligomeric state.

METHODS

Cloning and heterologous expression of GzmA. All DNA modifying enzymes were purchased from New England Biolabs or Stratagene and were used according to manufacturer's guidelines. Protein assay Bradford reagent was purchased from Bio-Rad and used according to manufacturer's guidelines. The *Pichia pastoris* expression system was purchased from Invitrogen.

The 705-base pair cDNA encoding mature human granzyme A, and the 693 -base pair cDNA encoding mature mouse granzyme A were subcloned into the yeast vector, pPICzαA (Invitrogen). Transformation into the *Pichia pastoris* strain and expression followed the manufacturer's guidelines.

QuikChange™ (Stratagene) site directed mutagenesis reaction was performed on the human GzmA gene to remove the N-linked glycosylation site, Asn153. The H→M GzmA was constructed via Quikchange with a deletion at position 217 and the following mutations: E218G, N219G and G226W. Expression and purification of all mutants was identical to the wild-type enzymes.

Purification of recombinant GzmA. After 48 h of induction with methanol, the supernatant from the GzmA expressing culture was harvested. The supernatant was adjusted to 50 mM NaCl and loaded onto a 100 ml SP-sepharose cation exchange column (Amersham Biosciences). The column was washed with five column volumes of 50 mM 2-(N-Morpholino)-ethanesulfonate (MES), pH 6.0 and 50 mM NaCl and eluted with a linear salt gradient of 50 mM to 1000 mM NaCl. Active GzmA eluted at 400 mM NaCl. The fractions from the SP-sepharose column were pooled and dialyzed against 50 mM MES pH 6.0, 100 mM NaCl and loaded onto a 1 ml Mono-S cation exchange column (Amersham Biosciences). The Mono-S column was washed with eight column volumes of buffer containing 50 mM MES pH 6.0, 100 mM NaCl. GzmA was eluted with a salt gradient from 100 mM to 800 mM. The final product was ≥98% pure as judged by SDS-PAGE Coomassie brilliant blue staining. The final yield of mouse GzmA was ~0.5 mg l^{-1} and that of human GzmA was ~10 mg l^{-1}.

The concentration of mouse and human GzmA protein was determined by absorbance measured at 280 nm and based on an extinction coefficient of 31520 M^{-1} cm^{-1} (1.22 ml mg^{-1})[40].

Positional scanning combinatorial library screening. Preparation and screening of the positional scanning synthetic combinatorial libraries were carried out as previously described[19]. Briefly, the concentration of each of the 6,859 substrates per well in the P1-Diverse library was 0.013 μM and that for the 361 substrates per well in the P1-Arg library was 0.25 μM. GzmA concentration was approximately 10–100 nM. Hydrolysis of the substrates was monitored fluorometrically for release of the ACC leaving group at an excitation wavelength of 380 nm and emission wavelength of 450 nm in 50 mM HEPES pH 7.4, 200 mM NaCl and 0.01% (v/v) Tween-20 at 30 °C.

Single substrate kinetic analysis. GzmA hydrolytic activity of ACC substrates was monitored at an excitation wavelength of 380 nm and an emission wavelength of 450 nm at 30 °C in a buffer containing 50 mM HEPES, pH 7.4, 200 mM NaCl and 0.01% (v/v) Tween-20. The final concentration of purified Ac-VANR-ACC substrate ranged from 0.005 to 1 mM. Because of the decreased solubility of Ac-GYFR-ACC at high concentrations, the final concentration of this substrate ranged from 1–50 μM. Initial rates were fit to the Michaelis-Menten equation describing steady state enzyme kinetics.

Crystallization. N153Q human GzmA was dialyzed into 50 mM MES pH 6.0, 50 mM NaCl and concentrated to 10 mg ml^{-1}. Enzyme was incubated overnight at 4 °C with 2.9 mM D-Phe-Pro-Arg-CMK. Crystals were grown at 17 °C by hanging drop diffusion with a reservoir solution of 0.1 M Tris pH 8.5, 0.2 mM Li_2SO_4, 13–18% (v/v) PEG 4K. The crystals were of the space group $C222_1$ with a unit cell of $a = 115.03$ Å, $b = 145.02$ Å, $c = 39.56$ Å.

Data collection and structure determination. Crystals were soaked in 0.1 M Tris, pH 8.5, 0.2 M Li_2SO_4, 15% (w/v) PEG 4K with 15% (v/v) PEG 400 as a cryoprotectant. Data were collected at 100 K at beamline 8.3.1 of the Advanced Light Source, Lawrence Livermore National Laboratories. Diffraction images were indexed and integrated with the DENZO[41] software and then scaled and merged with SCALEPACK[41]. The phases were determined by molecular replacement (EPMR[42]) using factor D (PDB entry 1DST) as the search model.

Structure refinement. Initial phases were calculated from the molecular replacement solution. Model bias was reduced using DM[43] followed by density modification in ARP/wARP[44]. The resulting electron density maps were traced in O[45]. The model was refined with refmac5 using the maximum likelihood target function[46]. Individual isotropic B-factors were refined after the R_{free} was below 30%. Explicit solvent was added to the model both by hand and using

ARP/wARP after the R_{free} was below 27%. All protein structure figures were generated using the program SPOCK (http://quorum.tamu.edu/spock/).

Construction of the mouse granzyme A homology model. The sequence of mouse GzmA was threaded onto the final structure of human GzmA via the SWISSModel server[47]. The most divergent active site loops (216–219 and 94–100) of the resulting structure were then minimized followed by 1ps of molecular dynamics in InsightII using the cvff forcefield and an explicit solvent model (Accelrys).

Coordinates. Coordinates for human N153Q GzmA have been deposited into the Protein Data Bank (accession code 1ORF).

ACKNOWLEDGMENTS

The authors would like to express thanks to J. Holton for technical assistance during data collection at the ALS beam line 8.3.1. We would also like to acknowledge the Macromolecular Structure Group at the University of California, San Francisco for computational support. This work was supported in part by a US National Institutes of Health post-doctoral training grant to J.K.B. and grants from the US National Institutes of Health to C.S.C. and R.J.F.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 23 April; accepted 23 May 2003
Published online 22 June 2003; doi: 10.1038/nsb944

1. Russell, J.H. & Ley, T.J. Lymphocyte-mediated cytotoxicity. *Annu. Rev. Immunol.* **20**, 323–370 (2002).
2. Kam, C.M., Hudig, D. & Powers, J.C. Granzymes (lymphocyte serine proteases): characterization with natural and synthetic substrates and inhibitors. *Biochem. Biophys. Acta* **1477**, 307–323 (2000).
3. Masson, D. & Tschopp, J. A family of serine esterase in lytic granules of cytolytic T lymphocytes. *Cell* **49**, 679–685 (1987).
4. Shresta, S., Graubert, T.A., Thomas, D.A., Raptis, S.Z. & Ley, T.J. Granzyme A initiates an alternative pathway for granule-mediated apoptosis. *Immunity* **10**, 595–605 (1999).
5. Beresford, P.J., Xia, Z., Greenberg, A.H. & Lieberman, J. Granzyme A loading induces rapid cytolysis and a novel form of DNA damage independently of caspase activation. *Immunity* **10**, 585–594 (1999).
6. Beresford, P.J. *et al.* Granzyme A activates an endoplasmic reticulum-associated caspase-independent nuclease to induce single-stranded DNA nicks. *J. Biol. Chem.* **276**, 43285–43293 (2001).
7. Fan, Z., Beresford, P.J., Oh, D.Y., Zhang, D. & Lieberman, J. Tumor suppressor NM23-H1 is a granzyme A-activated DNase during CTL-mediated apoptosis, and the nulceosome assembly protein SET is its inhibitor. *Cell* **112**, 659–672 (2003).
8. Irmler, M. *et al.* Granzyme A is an interleukin 1 β-converting enzyme. *J. Exp. Med.* **181**, 1917–1922 (1995).
9. Suidan, H.S. *et al.* Granzyme A released upon stimulation of cytotoxic T lymphocytes activates the thrombin receptor on neuronal cells and astroctyes. *Proc. Natl. Acad. Sci. USA* **91**, 8112–8116 (1994).
10. Zhang, D., Beresford, P.J., Greenberg, A.H. & Lieberman, J. Granzymes A and B directly cleave lamins and disrupt the nuclear lamina during granule-mediated cytolysis. *Proc. Natl. Acad. Sci. USA* **98**, 5746–5751 (2001).
11. Zhang, D. *et al.* Induction of rapid histone degradation by the cytotoxic T lymphocyte protease Granzyme A. *J. Biol. Chem.* **276**, 3683–3690 (2001).
12. Fan, Z., Beresford, P.J., Zhang, D. & Lieberman, J. HMG2 interacts with the nucleosome assembly protein SET and is a target of the cytotoxic T-lymphocyte protease granzyme A. *Mol. Cell. Biol.* **22**, 2810–2820 (2002).
13. Fan, Z. *et al.* Cleaving the oxidative repair protein Ape1 enhances cell death mediated by granzyme A. *Nat. Immunol.* **4**, 145–153 (2003).
14. Odake, S. *et al.* Human and murine cytotoxic T lymphocyte serine proteases subsite mapping with peptide thioester substrates and inhibition of enzyme activity and cytolysis by isocoumarins. *Biochemistry* **30**, 2217–2227 (1991).
15. Pasternack, M.S., Bleier, K.J. & McInerney, T.N. Granzyme A binding to target cell proteins. Granzyme A binds to and cleaves nucleolin in vitro. *J. Biol. Chem.* **266**, 14703–14708 (1991).
16. Masson, D., Zamai, M. & Tschopp, J. Identification of granzyme A isolated from cytotoxic T-lymphocyte-granules as one of the proteases encoded by CTL-specific genes. *FEBS Lett.* **208**, 84–88 (1986).
17. Schechter, I. & Berger, A. On the active site of protease. 3. Mapping the active of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Comm.* **32**, 898–902 (1968).
18. Thornberry, N.A. *et al.* A combinatorial approach defines specificity of members of the caspase family and granzyme B. Functional relationships estabilshed for key mediators of apoptosis. *J. Biol. Chem.* **272**, 17907–17911 (1997).
19. Harris, J.L. *et al.* Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. USA* **97**, 7754–7759 (2000).
20. Backes, B.J., Harris, J.L., Leonetti, F., Craik, C.S. & Ellman, J.A. Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin. *Nat. Biotechnol.* **18**, 187–193 (2000).
21. Kim, S., Narayana, S.V. & Volanakis, J.V. Crystal structure of a complement factor D mutant expressing enhanced catalytic activity. *J. Biol. Chem.* **270**, 24399–24405 (1995).
22. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
23. Galvin, J.P. *et al.* Apoptosis induced by granzyme B-glycosaminoglycan complexes: implications for granule-mediated apoptosis in vivo. *J. Immunol.* **164**, 5345–5350 (1999).
24. Poe, M. *et al.* Human cytotoxic lymphocyte tryptase. Its purifcation from granules and the characterization of inhibitor and substrate specificity. *J. Biol. Chem.* **263**, 13215–13222 (1988).
25. Barrios, A.M. & Craik, C.S. Scanning the prime-site substrate specificity of proteolytic enzymes: A novel assay based on ligand-enhanced lanthanide ion fluorescence. *Bioorg. Med. Chem. Lett.* **12**, 3619–3623 (2002).
26. Simon, M.M., Hoschutzky, H., Fruth, U., Simon, H.G. & Kramer, M.D. Purification and characterization of a T cell specific serine proteinase (TSP-1) from cloned cytolytic T lymphocytes. *EMBO J.* **5**, 3267–3274 (1986).
27. Fruth, U. *et al.* A novel serine proteinase (HuTSP) isolated from a cloned human CD8+ cytolytic T cell line is expressed and secreted by activated CD4+ and CD8+ lymphocytes. *Eur. J. Immunol.* **17**, 1625–1633 (1987).
28. Krahenbuhl, O. *et al.* Characterization of granzymes A and B isolated from granules of cloned human cytotoxic T lymphocytes. *J. Immunol.* **141**, 3471–3477 (1988).
29. Murphy, M.E. *et al.* Comparative molecular model building of two serine proteinases from cytotoxic T lymphocytes. *Proteins* **4**, 190–204 (1988).
30. Harris, J.L., Peterson, E.P., Hudig, D., Thornberry, N.A. & Craik, C.S. Definition and redesign of the extended substrate specificity of granzyme B. *J. Biol. Chem.* **273**, 27364–27373 (1998).
31. Waugh, S.M., Harris, J.L., Fletterick, R. & Craik, C.S. The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity. *Nat. Struct. Biol.* **7**, 762–765 (2000).
32. Earnshaw, W.C., Martins, L.M. & Kaufmann, S.H. Mammalian caspases: Structure, activation, substratea and functions during apoptosis. *Annu. Rev. Biochem.* **68**, 383–424 (1999).
33. Pereira, P.J. *et al.* Human β-tryptase is a ring-like tetramer with active sites facing a central pore. *Nature* **392**, 306–311 (1998).
34. Perona, J.J., Tsu, C.A., Craik, C.S. & Fletterick, R.J. Crystal structure of an ecotin-collagenase complex suggest a model for recognition and cleavage of the collagen triple helix. *Biochemistry* **36**, 5381–5392 (1997).
35. Eggers, C.T., Wang, S.X., Fletterick, R. & Craik, C.S. The role of ecotin dimerization in protease inhibition. *J. Mol. Biol.* **308**, 975–991 (2001).
36. Wang, S.X., Esmon, C.T. & Fletterick, R.J. Crystal structure of thrombin-ecotin reveals conformational changes and extended interactions. *Biochemistry* **40**, 10038–10046 (2001).
37. Tsuzuki, S. *et al.* Purification and identification of a binding protein for pancreatic secretory trypsin inhbitor: a novel role of the inhibitor as an anti-granzyme A. *Biochem. J.* **372**, 227–233 (2003).
38. Hink-Schauer, C., Estebanez-Perpina, E., Bode, W. & Jenne, D.E. Crystal structure of the apoptosis-inducing human granzyme A dimer. *Nat. Struct. Biol.* **10**, 535–540 (2003).
39. Jackson, D.S. *et al.* Synthesis and evaluation of diphenyl phosphonate esters as inhibitors of the trypsin-like granzymes A and K and mast cell tryptase. *J. Med. Chem.* **41**, 2289–2301 (1998).
40. Gill, S.C. & von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326 (1989).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
42. Kissinger, C.R., Gehlhaar, D.K. & Fogel, D.B. Rapid Automated Molecular Replacement By Evolutionary Search. *Acta Crystallogr. D* **55**, 484–491 (1999).
43. Cowtan, K. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4/ESF-EACBM Newsletter on Prot. Crystallogr.* **31**, 34–38 (1994).
44. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta Crystallogr. D* **49**, 129–149 (1993).
45. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
46. Murshudov, G., Vagin, A. & Dodson, E. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
47. Peitsch, M.C. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* **24**, 274–279 (1996).