Long-range Electrostatic Complementarity Governs Substrate Recognition by Human Chymotrypsin C, a Key Regulator of Digestive Enzyme Activation*§

Received for publication, January 29, 2013, and in revised form, February 15, 2013 Published, JBC Papers in Press, February 19, 2013, DOI 10.1074/jbc.M113.457382

Jyotica Batra†1, Andráš Szabó†1, Thomas R. Caulfield†‡, Alexei S. Soares‡, Miklós Sahin-Tóth§2, and Evette S. Radisky†3

From the Departments of †Cancer Biology and §Neuroscience, Mayo Clinic Cancer Center, Jacksonville, Florida 32224, the ‡Department of Molecular and Cell Biology, Boston University Henry M. Goldman School of Dental Medicine, Boston, Massachusetts 02118, and the ¶Biology Department, Brookhaven National Laboratory, Upton, New York 11973

Background: Chymotrypsin C (CTRC) targets specific regulatory cleavage sites within trypsinogens and procarboxypeptidases.

Results: The crystal structure of CTRC reveals the structural basis of substrate specificity.

Conclusion: Long-range electrostatic and hydrophobic complementarity drives CTRC association with preferred substrates.

Significance: The observations reveal the mechanistic basis for CTRC selectivity in digestive enzyme activation and degradation.

Intersubstrate recognition by digestive proteases is critical for optimal digestion and absorption of nutrients in the small intestine. Human chymotrypsin C (CTRC), a serine protease encoded by the PRTM12 gene, is a key regulator of digestive enzyme activation and degradation. CTRC is synthesized as an inactive zymogen that becomes active upon cleavage of specific regulatory sites within trypsinogens and procarboxypeptidases.

CTRC is capable of activating all three trypsinogens (1,2), yielding trypsin, a highly active serine protease that cleaves a broad range of proteins and peptides. CTRC is a member of the chymotrypsin family of proteases that share a common mechanism of catalysis involving the triad of catalytic residues Ser-195, His-57, and Asp-102 (3). The three functional domains of CTRC (4) are: the N-terminal domain (residues 1–128), the central catalytic triad domain (129–234), and the C-terminal domain (235–299). The catalytic triad domain contains the catalytic residues Ser-195, His-57, and Asp-102, which are involved in the catalytic mechanism for nucleophilic attack on the peptide bond during cleavage.

In contrast to trypsin, CTRC possesses significant binding affinity for procarboxypeptidases (5). CTRC is released by the pancreas and can activate procarboxypeptidases in the duodenum, where enteropeptidase initiates an activation cascade by specifically activating trypsinogen, which in turn activate chymotrypsinogen, proelastase, and procarboxypeptidase through a cascade of sequential activations.

Significance: The interaction between CTRC and substrates is governed by a combination of electrostatic and hydrophobic interactions. The crystal structure of CTRC reveals the structural basis of substrate specificity, providing insights into the mechanisms of substrate recognition and discrimination.

Conclusion: Long-range electrostatic and hydrophobic complementarity drives CTRC association with preferred substrates, enabling the selective activation of trypsinogens and procarboxypeptidases.

Background: CTRC is a key player in digestive enzyme activation, and its selective activation of trypsinogens and procarboxypeptidases is critical for optimal digestion and absorption of nutrients.

Results: The crystal structure of CTRC reveals the structural basis of substrate specificity, providing insights into the mechanisms of substrate recognition and discrimination.

Conclusion: The observations reveal the mechanistic basis for CTRC selectivity in digestive enzyme activation and degradation.

Significance: The interactions between CTRC and substrates are governed by a combination of electrostatic and hydrophobic interactions, providing insights into the mechanisms of substrate recognition and discrimination.
CTRC distinguishes identity with other human chymotrypsin isoforms, which typ-
ically cleave after aromatic P1 residues. CTRC has 50–66% sequence identity with human
pancreatic elastase isoforms, which have broad specificity for substrate residues
surrounding the cleavage site are numbered by the nomenclature of Schechter and Berger (70): starting from the scissile bond,
inability of other chymotrypsin and elastase isoforms to target trypsinogen
Ca2+-binding loop, it is apparent that the backbone conformational tendencies.

In the absence of a crystal structure for CTRC, our efforts to understand this specificity have made use of inhibitor phage
display selection and mutagenesis of the natural substrate cat-
itself from these other family members by uniquely targeting the regulatory
cleavage sites involved in trypsinogen activation (5) and degradation (6), and also serves as a co-activator, with trypsin, of procarboxypeptidases CPA1 and CPA2 (15). The inability of other chymotrypsin and elastase isoforms to target these regulatory sequences (5, 15, 16) points to the existence of unique elements of CTRC specificity, as does a recent study in which we have identified selective inhibitors of CTRC using phage display (17).

The mechanisms by which CTRC, confronted with multiple potential substrates and cleavage sites, selects from among them, is critical to understanding its protective role in the pancreas and its pathological role in disease. Given the opposing effects of trypsinogen cleavage by CTRC within the activation peptide versus the Ca2+-binding loop, it is apparent that the hierarchy of CTRC selectivity toward these competing sites has been carefully titrated by nature. Remarkably, even a very subtle mutation within a cleavage site can shift the activity of CTRC and tip the balance toward a disease state.

In the absence of a crystal structure for CTRC, our efforts to understand this specificity have made use of inhibitor phage
display selection and mutagenesis of the natural substrate cat-
onic trypsinogen, with somewhat contradictory results (16, 17). Inhibitor studies suggest strong preference for Leu at the P1 posi-
tion (17), but mutagenesis studies again show only moderate effects upon alteration of individual charged residues within the cat-
otinability of other chymotrypsin and elastase isoforms to target these regulatory sequences (5, 15, 16) points to the existence of unique elements of CTRC specificity, as does a recent study in which we have identified selective inhibitors of CTRC using phage display (17).

The mechanisms by which CTRC, confronted with multiple potential substrates and cleavage sites, selects from among them, is critical to understanding its protective role in the pancreas and its pathological role in disease. Given the opposing effects of trypsinogen cleavage by CTRC within the activation peptide versus the Ca2+-binding loop, it is apparent that the hierarchy of CTRC selectivity toward these competing sites has been carefully titrated by nature. Remarkably, even a very subtle mutation within a cleavage site can shift the activity of CTRC and tip the balance toward a disease state.

In the absence of a crystal structure for CTRC, our efforts to understand this specificity have made use of inhibitor phage
display selection and mutagenesis of the natural substrate cat-
onic trypsinogen, with somewhat contradictory results (16, 17). Inhibitor studies suggest strong preference for Leu at the P1 posi-
tion (17), but mutagenesis studies again show only moderate effects upon alteration of individual charged residues within the cat-
otinability of other chymotrypsin and elastase isoforms to target these regulatory sequences (5, 15, 16) points to the existence of unique elements of CTRC specificity, as does a recent study in which we have identified selective inhibitors of CTRC using phage display (17).

The mechanisms by which CTRC, confronted with multiple potential substrates and cleavage sites, selects from among them, is critical to understanding its protective role in the pancreas and its pathological role in disease. Given the opposing effects of trypsinogen cleavage by CTRC within the activation peptide versus the Ca2+-binding loop, it is apparent that the hierarchy of CTRC selectivity toward these competing sites has been carefully titrated by nature. Remarkably, even a very subtle mutation within a cleavage site can shift the activity of CTRC and tip the balance toward a disease state.

In the absence of a crystal structure for CTRC, our efforts to understand this specificity have made use of inhibitor phage
display selection and mutagenesis of the natural substrate cat-
onic trypsinogen, with somewhat contradictory results (16, 17). Inhibitor studies suggest strong preference for Leu at the P1 posi-
tion (17), but mutagenesis studies again show only moderate effects upon alteration of individual charged residues within the cat-

| Substrate cleavage site | P3 | P4 | P5 | P2 | P1 | P1’ | P2’ | P3’ | P4’ |
|-------------------------|----|----|----|----|----|-----|-----|-----|-----|
| **Tg1, Tg2 activation peptide** | -  | -  | A  | P  | F  | D   | D   | D   | K  |
| **Tg1 p.A16V activation peptide** | -  | -  | V  | P  | F  | D   | D   | D   | K  |
| proCPA1 activation peptide | E  | D  | V  | Q  | S  | L   | L   | D   | E   | Q  |
| proCPA2 activation peptide | E  | D  | V  | Q  | S  | L   | L   | D   | K   | N  |
| **Tg1, Tg2 Ca2+ binding loop** | H  | N  | I  | E  | V  | E   | G   | N   | E   | Q  |
| **Tg3 Ca2+ binding loop** | H  | N  | K  | V  | E  | G   | N   | E   | Q  |   |
| **Tg1, Tg2 alternative site** | P  | Y  | Q  | V  | S  | L   | N   | S   | G   | Y   | H  |
| **Tg2 alternative site (Y124(N))** | E  | N  | S  | V  | P  | Y   | Q   | V   | S   | L   | N  |
| **Tg2 alternative site (L290(H))** | F  | C  | V  | G  | F  | L   | E   | G   | K   | D   |   |
| **Tg3 alternative site** | G  | W  | G  | N  | T  | L   | S   | F   | G   | A   |   |
| **Tg3 alternative site (Y1560(G))** | G  | N  | T  | L  | F  |   | G   | A   | D   |   |   |

Inhibitor reactive site

| Pe | P3 | P5 | P4 | P2 | P1 | P1’ | P2’ | P3’ | P4’ |
|----|----|----|----|----|----|-----|-----|-----|-----|
| eglin c (L355-K366) | G  | S  | P  | V  | T  | L   | D   | L   | R   | N   | -   |
| SGPI-2 (L30-K31) | S  | A  | A  | C  | T  | L   | K   | A   | C   | P   | Q   | (17) |
| **SGPI-2 library selection consensus** | (A) | A/G | (C) | T/S | M/L/P/Y | (K) | A/C | P/Q | (17) | (E/D) | (Q) | (17) |

The abbreviations used in the table are: Tg1, human cationic trypsinogen; Tg2, human anionic trypsinogen; Tg3, human mesotrypsinogen; proCPA1, human procarboxypeptidase A1; proCPA2, human procarboxypeptidase A2; SGPI-2, Schistocerca gregaria proteinase inhibitor-2. ***, substrate peptides shaded in gray have been modeled in the CTRC active site as described in the text, structural coordinates for the models are provided as supplemental Models S1–S3. **, sequence shown is the consensus of multiple clones selected from a phage-displayed SGPI-2 library that was diversified at P4, P2, P1, P1’, P2’, P3’, P4’. Residues shown in parentheses were at positions that were not diversified in the library.

Located within the active site cleft between the two β-barrels (13, 14), CTRC has 50–66% sequence identity with human pancreatic elastase isoforms, which have broad specificity for cleavage after hydrophobic P1 residues, and ~40% sequence identity with other human chymotrypsin isoforms, which typically cleave after aromatic P1 residues. CTRC distinguishes itself from these other family members by uniquely targeting the regulatory cleavage sites involved in trypsinogen activation (5) and degradation (6), and also serves as a co-activator, with trypsin, of procarboxypeptidases CPA1 and CPA2 (15). The inability of other chymotrypsin and elastase isoforms to target these regulatory sequences (5, 15, 16) points to the existence of unique elements of CTRC specificity, as does a recent study in which we have identified selective inhibitors of CTRC using phage display (17).

The mechanisms by which CTRC, confronted with multiple potential substrates and cleavage sites, selects from among them, is critical to understanding its protective role in the pancreas and its pathological role in disease. Given the opposing effects of trypsinogen cleavage by CTRC within the activation peptide versus the Ca2+-binding loop, it is apparent that the hierarchy of CTRC selectivity toward these competing sites has been carefully titrated by nature. Remarkably, even a very subtle mutation within a cleavage site can shift the activity of CTRC and tip the balance toward a disease state.

In the absence of a crystal structure for CTRC, our efforts to understand this specificity have made use of inhibitor phage
display selection and mutagenesis of the natural substrate cat-
onic trypsinogen, with somewhat contradictory results (16, 17). Inhibitor studies suggest strong preference for Leu followed by Met at the P1 position (17), whereas substrate mutagenesis suggests that CTRC is much more permissive of alternative P1 residues (16). Another potential specificity feature is suggested by the presence of multiple Asp and Glu resi-
dues within favored CTRC target sequences (Table 1); these acidic residues appear consistently in the P4’ position but are otherwise non-uniformly situated. Inhibitor phage display confi-

Here, the first structure of active CTRC reveals a familiar fold with distinctive electrostatic features surrounding the substrate

cleavage cleft. The CTRC active site is occupied by the inhibitor eglin c bound in a substrate-like conformation, offering insights into the structural basis for the unique substrate specificity of CTRC toward key physiological and pathological substrate sequences. Analysis of the structure suggests that whereas the bulk of binding energy derives from burial of the P1 residue and several other hydrophobic side chains, specificity may derive largely from the exaggerated role of long-range electrostatic interactions, from a moderate preference for Leu at the P1 position, and may also be influenced by local sequence-dependent backbone conformational tendencies.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—Human chymotrypsino-
gen C bearing a C-terminal His10 tag was expressed using transiently transfected HEK 293T cells, purified using metal chelation chromatography, dialyzed against 50 mM sodium phosphate (pH 8.0) and 300 mM NaCl, concentrated to 20 mg/ml, and activated by cleavage with cationic trypsin as previously described (4).

A construct for recombinant bacterial expression of eglin c from *Hirudo medicinalis* was a generous gift from Professor Robert S. Fuller, University of Michigan Medical School. Eglin c was expressed in *Escherichia coli* host strain BL21(DE3) and purified to homogeneity from periplasmic extract by SP-Sep
Structure of the CTRC-Eglin c Complex

harose chromatography essentially as previously described (18). Eglin c was dialyzed into 10 mM (NH₄)₂OAc at pH 6.0, lyophilized, and stored at −80 °C until use, when it was reconstituted with H₂O.

Crystallization—A 1:1 (mol/mol) mixture of CTRC and eglin c was concentrated to achieve a protein concentration of 4–5 mg/ml. Crystallization employed the hanging drop vapor diffusion method. Diffraction quality crystals were grown at room temperature from droplets containing 0.2 M lithium sulfate monohydrate, 0.1 M Tris-HCl (pH 8.5), and 30% (w/v) PEG 4000. Crystals grew over the course of 3 weeks to 0.1 mm. Crystals were soaked in a cryoprotectant solution (0.2 M lithium sulfate monohydrate, 0.1 M Tris-HCl (pH 8.5), and 30% (w/v) PEG 4000, and 15% glycerol) and cryocooled in liquid N₂.

Crystals were screened for diffraction and data were collected at 100 K at beamline X29 at the National Synchrotron Light Source, Brookhaven National Laboratory. The best crystals diffracted to 1.9-Å resolution. We identified crystals of CTRC/eglin c belonging to orthorhombic space group P2₁2₁2₁, with unit cell dimensions a = 56.27, b = 76.25, c = 81.82, and containing one complex in the asymmetric unit. The data were merged and scaled using DENZO/SCALEPACK (19).

Structure Determination—The x-ray structure of chymotrypsin C in complex with eglin c was solved by molecular replacement using Phaser (20) supported by CCP4. The elastase chain from the previously solved structure of a porcine elastase-inhibitor complex (PDB code 1EAI, chain A) (21) and eglin c from bovine α-chymotrypsin-eglin c (PDB IACB, chain I) (22) were used as search models. ARP/wARP was employed after molecular replacement for automated rebuilding of the CTRC-eglin c complex structure (23, 24). A test set of 5% of the total reflections was excluded from rebuilding and refinement of the model. Refmac5 (25) was used to carry out refinement and to place water molecules into difference peaks (Fo – Fc) greater than 3σ; manual rebuilding was done using COOT (26).

A 10-residue activation peptide chain that was not liberated following proteolytic activation of CTRC due to a disulfide link with the enzyme (chain Q) was manually built into electron density maps (2Fo – Fc) using COOT. Phosphate ions were also added using COOT. The final stage of the restrained refinement included water molecules with peaks greater than 1σ and within acceptable H-bonding distances from neighboring protein atoms and three phosphate ions. Surprisingly, we were unable to build either the His₁₀ tag or the N-linked glycosyl groups of CTRC due to the absence of electron density. The final R/Rfree was 0.157/0.210. Figs. 1, 2, 4A, and 5 were generated using PyMOL (Schrödinger, LLC).

Homology Modeling of Human Elastase and Chymotrypsin Isoforms—Homology models of human elastases (ELA2A, ELA3A, and ELA3B) and chymotrypsins (CTR1B, CTR2B, and CTRL1) were constructed using the SWISS-MODEL workspace (27). Homology models were constructed using porcine elastase bound to an Ascaris chymotrypsin/elastase inhibitor (PDB code 1EAI, chain A) (21) or our model of active CTRC bound to eglin c (PDB code 4H4F, chains A and Q) as templates with 40–50% amino acid identity to the modeled proteins. Molecular surfaces were generated using the Molecular Surface module of Schrödinger 2012 (Schrödinger, LLC). Molecular surfaces were based upon high-resolution settings (0.3 Å grid, probe radius 1.4 Å, and van der Waals radius scale of 1.0 Å). All surfaces were rendered in red - white - blue color scale using PB electrostatic potential from calculated charges at pH 8.5 with the color ramp set to a minimum of −0.35 and maximum of 0.15 for the structures shown in Figs. 3 and 4; the figures were generated using the Maestro all-purpose molecular modeling environment, version 9.3.023 from the Schrödinger Suite 2012.

Molecular Modeling of CTRC Substrate Binding—The x-ray structure for CTRC was imported into the Protein-Preparation-Wizard graphical user interface of Schrödinger with Maestro 2012 version 9.3.5 (Schrödinger, LLC) for adaption to the OPLS2005 force field. Bond orders were assigned, zero-order bonds to metals were determined, disulfide bonds were created, and all hydrogens were generated for every residue. Hydrogen bond assignment was based on sampling water orientations and taking into account crystallographic waters. Protonation states were predicted for pH 8.5 using PROPKA (28, 29). Steric clashes were resolved with convergence of a root mean square deviation to 0.3 Å using the OPLS2005 force field within the Schrödinger interface.

For modeling and molecular docking of peptide substrate sequences, starting conformations of substrates were obtained by Polak-Ribiere Conjugate Gradient energy minimization with the OPLS 2005 force field for 5000 steps, or until the energy difference between subsequent structures was <0.001 kJ/mol Å (28). Force field minimization used a water-based solvent, generating charges with an extended cutoff (van der Waals 8.0 Å, electrostatic 20 Å, H-bond 4.0 Å). We placed soft restraints on all residues >6 Å from the modeled substrate by using harmonic restraints at 100 kcal/mol, and allowed the residues within the 6-Å cutoff to move freely during Polak-Ribiere Conjugate Gradient energy minimization over 500 iterations with repetition as necessary to converge upon a gradient threshold of <0.05.

We have previously described the methodology used for substrate docking (31); briefly, the binding site was generated via overlapping grids based on the x-ray structure with a default rectangular box centered on the target substrate. Substrates were docked into the binding site of CTRC using Glide extra precision (XP) (Glide, version 5.6, Schrödinger, LLC); molecular conformations were sampled using methods described previously (32). A structure-based pharmacophore score was generated from the optimized, best scoring pose for each substrate ligand based on the descriptors from the Glide XP score using an established approach (31, 33, 34). The energetic value assigned to each pharmacophore feature was calculated using Phase (Phase, version 3.2, Schrödinger, LLC) as the sum of the Glide XP contributions of the atoms comprising the site. Overall dockings at the active site were quantified and ranked on the basis of these energetic terms (33, 34). To account for protein flexibility and lessen the effects of minor steric clashes, excluded volumes spheres corresponding to 80% of the van der Waals atomic radii were created for all CTRC atoms within 6 Å of each substrate or mutagenized residue modeled. A minimum of two poses per substrate, chosen for a combination of best-
TABLE 2
Crystallographic data collection and refinement statistics for CTRC-eglin c complex

| PDB ID | 4H4F |
|--------|------|
| Complexes per ASU | 1 |
| Space group | P212121 |
| Unit cell, Å | 56.27, 76.25, 81.82 |
| Resolution, Å | 1.9 |
| Unique reflections | 27,661 |
| Completeness (%) | 97.34 (76.8) |
| Multiplicity | 6.8 (4.8) |
| Mean I/σ (σ) | 51.5 (9.0) |
| Rmerge | 0.043 (0.146) |
| Rmerge/Rfree (%) | 0.157 (0.210) |
| Root mean square deviation bonds, Å | 0.072 |
| Root mean square deviation angles (degree) | 2.066 |
| Protein atoms | 2469 |
| Ions | 3 (PO4) |
| Water molecules | 320 |
| Φ/Ψ angle distribution | 209 (87.8%) |
| In additionally allowed regions | 28 (11.8%) |
| In generously allowed regions | 1 (0.4%) |

Values in parentheses are for the highest resolution shell (1.93–1.9 Å).
Ramachandran distribution is reported as defined by PROCHECK/PDB validation.

crystallographic data was collected and refined to a final 1.9 Å resolution in space group P212121 with one bimolecular complex per asymmetric unit. The structure was solved by molecular replacement using as search models the previously reported structures of porcine pancreatic elastase (PDB 1EAI) (21), featuring 52% sequence identity to human CTRC, and eglin c (PDB 1ACB) (22). A structure of the bovine CTRC precursor chymotrypsinogen C, possessing 80% sequence identity to human CTRC, has been reported previously (PDB 1PYT) (35); however, the elastase structure was judged to offer a better model as a result of substantial conformational changes (36). The structure was rebuilt and refined to a final Rcryst (Rfree) of 15.7% (21.0%); crystallographic statistics are summarized in Table 2.

Overall Structure of the CTRC-Eglin c Complex—Like other serine proteases of the chymotrypsin family (37), CTRC is comprised of two β-barrels, at the interface of which is located the active site containing the catalytic triad of Ser195 (Ser216),6 His57 (His74), and Asp102 (Asp121) (Fig. 1A). The substrate binding cleft between the β-barrels is occupied by bound inhibitor eglin c, a 70-amino acid protease inhibitor originally isolated from the leech H. medicinalis (38). As has been described previously for eglin c (22, 39, 40) and for the structurally related chymotrypsin inhibitor 2 (41, 42), the inhibitor is a wedge-shaped molecule featuring a hydrophobic core formed by a helix and a small β-sheet, from which protrudes the inhibitory canonical loop, forming the thin edge of the wedge, which fits into the substrate binding cleft of the enzyme (Fig. 1, A and B). The canonical loop of eglin c, comprised of residues 40–50, binds to the active site of CTRC in the substrate-like fashion typical of canonical serine protease inhibitors (43–45) (Fig. 1B).

CTRC residues 1–10 of the cleaved activation peptide chain (CGVPVSFPNFL) are retained by the activated enzyme due to a disulfide link between Cys4 (Cys17) of the activation peptide and Cys12 (Cys141), located in the linker between the two β-barrels on the enzyme face distal from the active site (Fig. 1C). The disulfide bonding pattern and consequent retention of the activation peptide, conserved among other chymotrypsins and the elastase 2A isoform, has been demonstrated to stabilize the enzyme against denaturation and proteolytic digestion by pepsin (46). In addition to the covalent link, the activation peptide association is stabilized by two backbone-backbone H-bonds of Gly2 (Gly18), an H-bond between the carboxyl of Pro7 (Pro23) and the side chain of Arg23 (Arg27), and substantial hydrophobic interactions of Pro4 (Pro20), Phe6 (Phe22), Pro8 (Pro24), and Leu19 (Leu26) (Fig. 1C). Clear density is observed for the C-terminal Leu19 (Leu26) of the activation peptide chain, revealing that residues Ser11–Ala12–Arg15 (Ser27–Ala31–Arg36) of the activation peptide chain have been proteolytically removed. Given the binding preference of CTRC for Leu at the P1 position (17), it is probable that removal of the tripeptide is accomplished through autoproteolytic cleavage in trans, as suggested previously (47).

Although the recombinant CTRC possessed a His10 tag, none of the residues of the tag were visible in the electron density, suggesting that either the tag was disordered or had been proteolytically removed post-purification by autoproteolysis. Likewise, although CTRC has been shown to be glycosylated on Asn368 (Asn52), a modification required for efficient folding and secretion (48), we did not observe density for the glycosyl group. The side chain of Asn368 (Asn52) was poorly defined and it is likely that the site at this site was present but disordered.

Insights into Chymotrypsinogen C Activation—Comparison of the human CTRC structure with the earlier reported structure of bovine chymotrypsinogen C (PDB 1PYT) (35) shows that upon cleavage of the activation loop at Arg15–Val16 (Arg23–Val26), the new N-terminal residue Val16 (Val23) becomes buried, forming a salt bridge with Asp194 (Asp215) and an H-bond with the carbonyl oxygen of Arg13 (Arg16). This reorganization has little impact on the structure of the first β-barrel domain, but results in major conformational alterations of the second β-barrel domain, particularly within loops 1 and 2, which shape the S1 specificity pocket and oxyanion hole, and loop D, which helps to shape the primed side subsites. Although the catalytic triad residues are already roughly positioned in the second β-barrel domain, particularly within loops 1 and 2, which shape the S1 specificity pocket and oxyanion hole, and loop D, which helps to shape the primed side subsites.

Details of the Inhibitory Interaction—The inhibitory loop of eglin c is stabilized in the distinctive canonical conformation by

---

6The CTRC residue numbering used in this report and in the crystal structure coordinates is derived by homology to bovine chymotrypsin, the archetype member of this peptidase family, for consistency with structural literature in the serine protease field. Designations in parentheses are the corresponding residue numbers based on sequential numbering of the CTRC precursor.
a hydrophobic “mini-core” (49) comprised of Leu37, Val43, and Phe55, and by a H-bond network involving Thr44, Asp46, the Arg48 backbone N, Arg51, Arg53, and the C-terminal carboxyl group of Gly70 (Fig. 2A). This H-bond network is critical for maintaining protease affinity and resistance to proteolysis of the inhibitor (50–54).

The mode of protease binding of canonical inhibitors like eglin c has been shown to very closely mimic the reactive Michaelis complex with a true peptide substrate (45). As is typical of such complexes, the Leu45-Asp46 reactive site peptide bond of eglin c is appropriately positioned for nucleophilic attack by the catalytic Ser195 (Ser216) of CTRC (Fig. 2B), the initial step in enzyme-catalyzed proteolysis. The largely hydrophobic substrate binding cleft of CTRC is fully occupied by eglin c residues 40–50, which mimic non-primed side substrate residues P1–P6 and primed side residues P1–P5. The non-primed side residues are positioned by antiparallel backbone-backbone H-bonds between the P3 residue Val43 and CTRC Gly216 (Gly238), and between the P1 residue Leu45 and CTRC Ser214 (Ser236) (Fig. 2C). The Leu45 carbonyl oxygen is positioned to interact with the CTRC oxyanion hole amide nitrogens of Ser195 (Ser216) and Gly193 (Gly214). On the primed side, additional H-bonds are formed between P2/H2/P1/P5/P6 and the hydrophobic P2/H1/P1/P5/P6.

The S1 primary specificity pocket of CTRC, occupied in the complex by the side chain of eglin c Leu45, is shaped by the hydrophobic side chains of CTRC residues Ala190 (Ala211), Val213 (Val235), and Val226 (Val250) (Fig. 2C). A more shallow hydrophobic depression in the substrate binding cleft is shaped by the side chains of CTRC Leu99 (Leu118) and Phe215 (Phe237), which form a binding pocket for P4 residue Pro42 (Fig. 2C).

The backbone conformation of eglin c residues 43–49 bound to CTRC closely parallels that seen in the previously reported structure of eglin c with bovine α-chymotrypsin (PDB 1ACB (22)). By contrast, eglin c residues 39–42 are shifted compared with the bovine α-chymotrypsin-eglin c complex to lie 3 Å further removed from a basic patch formed by CTRC residues Arg175 (Arg195), Arg218 (Arg241), and Lys224 (Lys248) (Fig. 2D). This basic pocket forms the S6 subsite of CTRC, which is occupied in the complex by a coordinated phosphate ion in the absence of a side chain on the eglin c P6 residue Gly40 (Fig. 2D).

Structural Insights into CTRC Substrate Specificity—CTRC has been shown to act as a regulator of pancreatic zymogen activation by targeting a specific set of substrate cleavage sites not recognized by other chymotrypsin or elastase-like digestive proteases (Table 1) (5, 6, 15, 16). One element of this specificity is more highly efficient cleavage after Leu residues compared with other chymotrypsin and elastase isoforms (16). By contrast, the elastase isoforms show broad P1 specificity but comparatively low catalytic efficiency on short peptide substrates, whereas chymotrypsins prefer aromatic residues Phe, Tyr, or Trp at P1 (16). The position occupied in CTRC by Val226 (Val250), one of the hydrophobic residues shaping the S1 pocket (Fig. 2C), is in other chymotrypsins filled by Gly or Ala, and the
the S1 subsite of CTRC is still capable of accommodating aro-
us for cleavage of tetrapeptide substrates (16). Nevertheless,
play selection and inhibitor binding studies (17) and by
surrounding the substrate binding cleft of CTRC (Fig. 3,
particularly striking concentration of positive charge in a ring

target the same regulatory cleavage sites (Fig. 3). We observed a
other human chymotrypsin and elastase isoforms, which do not
homology models and calculated electrostatic surfaces for
of the CTRC structure, and for comparison we generated
ficity, we calculated the predicted electrostatic surface potential
side of the cleavage site (Table 1). To gain insight into the prob-
ematic residues, and in fact, CTRC catalytic rates are slightly

Another distinctive feature shared by the natural target sites
of CTRC is an unusual clustering of acidic residues. Asp or Glu
appear very frequently at the P4
residue Leu47 fills a pocket formed by CTRC Arg143 and Ile151. Multiple backbone H-bonds orient the inhibitor, indicated by black
dotted lines. D, positively charged P5 pocket displaces eglin c backbone to bind phosphate. Basic side chains of CTRC Arg72, Arg218, and Lys225 coordinate a
phosphate ion, displacing eglin c from the orientation in which it is found in complex with bovine α-chymotrypsin (shown in semitransparent white stick
representation; PDB code 1ACB). The 2Fo − Fc electron density map shown for the phosphate ion is contoured at 1.6σ.

The most intense concentrations of positive charge are created
by one cluster of basic residues on the non-primed side of the
cleft in the region that makes contact with P5 and P6 substrate
residues, and a second cluster of basic residues on the primed
side bordering the subsites that recognize P2

 Modeling of CTRC-Substrate Complexes—Selection by CTRC
among several potential cleavage sites within human cationic
trypsinogen has profound health implications. Whereas cleav-
age in the trypsinogen Ca2
-binding loop is a normal mech-
nism for CTRC to apply the brakes to a premature cascade of
activation within the trypsinogen Ca2
ase and chymotrypsin isoforms (Fig. 3, lower panels). The
human elastases feature small patches of both positive and neg-
ative charge in roughly equal proportions, whereas human chy-
motrypsins possess substrate binding clefts lined with primarily
negative charge. Thus, the unusual concentration of positive
charge surrounding the CTRC substrate binding cleft is likely to
be a major determinant driving specificity for substrate
sequences of net negative charge.

comparing to the orientation in which it is found in complex with bovine α-chymotrypsin (shown in semitransparent white stick representation; PDB code 1ACB). The 2Fo − Fc electron density map shown for the phosphate ion is contoured at 1.6σ.

FIGURE 2. Eglin c inhibitory interaction with CTRC. A, stabilization of the inhibitory loop of eglin c. The eglin c binding loop assumes a substrate-like canonical conformation of the peptide backbone, stabilized on the nonprimed side by hydrophobic interactions of Leu37, Val43, and Phe55, and on the primed side by an H-bond network involving Thr44, Asp56, the Arg58 amide nitrogen, Arg51, Arg53, and C-terminal Gly70. The side chain of Arg58 is omitted for clarity. B, the CTRC-eglin c complex resembles an enzyme-substrate Michaelis complex. The Leu45-Asp46 reactive site peptide bond of eglin c, linking the P1 and P1′ residues, lies in proper orientation for attack by the catalytic Ser195 of CTRC. The 2Fo − Fc electron density map is shown contoured at 2.0σ. C, key CTRC-eglin c binding interactions. The eglin c P1 residue Leu45 fills the S1 pocket bordered by CTRC Ala190, Val213, and Val226. P4 residue Pro42 fills a hydrophobic concavity formed by CTRC Leu99 and Phe215. P2 residue Leu47 fills a pocket formed by CTRC Arg143 and Ile151. Multiple backbone H-bonds orient the inhibitor, indicated by black dotted lines. D, positively charged P5 pocket displaces eglin c backbone to bind phosphate. Basic side chains of CTRC Arg72, Arg218, and Lys225 coordinate a phosphate ion, displacing eglin c from the orientation in which it is found in complex with bovine α-chymotrypsin (shown in semitransparent white stick representation; PDB code 1ACB). The 2Fo − Fc electron density map shown for the phosphate ion is contoured at 1.6σ.
site in our previously reported structure of human cationic trypsin (55), and we also used the new CTRC structure as a starting point to generate models of CTRC bound to specific substrate sequences.

In the cationic trypsin structure, Ca\textsuperscript{2+}/H11001 is coordinated within the loop by the side chains of residues Glu\textsuperscript{75} and Glu\textsuperscript{85}, which anchor the base of the loop, and by carbonyl oxygens of Asn\textsuperscript{77} and Val\textsuperscript{80} (Fig. 4A).\textsuperscript{7} CTRC targets the Leu\textsuperscript{81}-Glu\textsuperscript{82} peptide bond for cleavage. With Ca\textsuperscript{2+} bound, Leu\textsuperscript{81} is exposed and accessible to CTRC; however, the peptide backbone stretching away from the cleavage site is unable to assume the canonical conformation required for cleavage (56). Cationic trypsinogen possesses a large number of acidic residues in this region; in

\textsuperscript{7} The human cationic trypsinogen residue numbering used here is based on sequential numbering of the trypsinogen precursor as is conventionally used in designating natural polymorphisms, e.g. p.A16V.
addition to Glu\textsuperscript{75} and Glu\textsuperscript{85}, there are two additional glutamates in the loop (Glu\textsuperscript{79} and Glu\textsuperscript{82}), as well as three acidic residues in neighboring loops (Glu\textsuperscript{75}, Asp\textsuperscript{156}, and Glu\textsuperscript{157}; Fig. 4A). Thus, even with Ca\textsuperscript{2+} present to neutralize the charge of Glu\textsuperscript{75} and Glu\textsuperscript{85}, there is a strongly negative electrostatic potential covering this region of the molecule (Fig. 4B). This potential will generate macroscopic electrostatic complementarity between the Ca\textsuperscript{2+}-binding loop of cationic trypsin or trypsinogen and the substrate-binding site of CTRC. It is anticipated that when Ca\textsuperscript{2+} is released, this loop becomes more flexible and accessible to CTRC, able to assume a productive binding orientation, and that electrostatic attraction may increase further due to exposure of Glu\textsuperscript{75} and Glu\textsuperscript{85}.

Because eglin c binds to CTRC in the canonical orientation of an ideal substrate (45), we were able to model substrate sequences into the binding cleft of CTRC using eglin c as a template and then to optimize the interactions through energy minimization. Models were generated for CTRC bound to the human cationic trypsinogen activation peptide (APFDDDDK) and the CTRC-labile site within the cationic trypsinogen Ca\textsuperscript{2+}-binding loop (HNIIEVLEGNEQ). The resulting “global” total enthalpies (\(\Delta H\)) from energy minimization were \(-70,572\) and \(-71,057\) kcal/mol, respectively. Following modeling and minimization, we conducted substrate docking for each substrate with CTRC, giving docking scores of \(-10.16\) and \(-17.72\) kcal/mol, respectively. The cationic trypsinogen Ca\textsuperscript{2+}-binding loop, which contacts the greater number of non-primed side subsites, was predicted to be the more preferred substrate based on overall lowest energy from docking/-binding with CTRC.

The docked model of preferred substrate HNIIEVLEGNEQ shows the positioning of substrate residues relative to the electrostatic features of the CTRC surface (Fig. 4C); residues Asn\textsuperscript{77/-Ile\textsuperscript{78}, Glu\textsuperscript{79/-Val\textsuperscript{80}, Leu\textsuperscript{81/-Glu\textsuperscript{82}, Glu\textsuperscript{83/-Asn\textsuperscript{84}, Glu\textsuperscript{85}}}} fill the largely hydrophobic cleft between the flanking clusters of positive charge, with the Leu\textsuperscript{81} side chain embedded in the hydrophobic S\textsubscript{1} subsite. Surprisingly, none of the acidic side chains of the substrate, Glu\textsuperscript{75}, Glu\textsuperscript{82}, or Glu\textsuperscript{85}, which fill the P\textsubscript{3}, P\textsubscript{1}, and P\textsubscript{4} positions, respectively, form direct salt bridges with the clustered basic side chains of CTRC in the energy minimized docked model (Fig. 5A). Instead, it would appear that the complex stabilization attributable to charge complementarity derives from longer range electrostatic interactions. Trypsinogen Glu\textsuperscript{85} in the P\textsubscript{4} position is stabilized by CTRC Arg\textsuperscript{143}A via an interaction bridged by H-bonds with the P\textsubscript{4}‘ side chain of trypsinogen Asn\textsuperscript{84}. The trypsinogen Glu\textsuperscript{85} P\textsubscript{4}‘ side chain is positioned equidistant from the guanidinium groups of CTRC Arg\textsuperscript{39} and Arg\textsuperscript{143} (about 6 Å from each). The major favorable close interactions in the complex are comprised primarily of hydrophobic and van der Waals interactions, and a series of hydrogen bonds tethering the substrate peptide backbone within the binding cleft (Fig. 5A).

The docked model of the competing substrate cleavage site, APFDDDDK (the trypsinogen activation peptide), shows that this shorter substrate fills only the S\textsubscript{1}-S\textsubscript{3} subsites on the non-primed side of the cleft, forming fewer H-bonds and hydrophobic interactions as compared with the longer substrate (Fig. 5B). In this model, Phe\textsuperscript{18} fills the hydrophobic S\textsubscript{1} subsite, and the substrate does form a single direct salt bridge between Asp\textsuperscript{20} at the P\textsubscript{2}‘ position and Arg\textsuperscript{143} (Arg\textsuperscript{162}) of CTRC (Fig. 5B). However, as was the case for the Ca\textsuperscript{2+}-binding loop site, electrostatic stabilization conferred by the Asp residues at the P\textsubscript{1}, P\textsubscript{3}, and P\textsubscript{4}‘ positions is apparently mediated through longer-range interactions with CTRC.

Mutation of a single residue in the trypsinogen activation peptide, where Val is substituted for Ala\textsuperscript{16} at the P\textsubscript{3} position, predisposes carriers for development of pancreatitis, apparently by altering the balance of the CTRC substrate selectivity in favor of the activation peptide (4, 7, 8). To explore the structural basis for this shift in selectivity, we modeled the complex of CTRC with the p.A16V mutant sequence. The docked model of the p.A16V mutant trypsinogen activation peptide VPFD-
**Structure of the CTRC-Eglin c Complex**

DDDK very closely resembles that of the wild-type sequence, with significant differences confined to the mutated residue itself. The N-terminal amine of Ala196 H-bonds to the CTRC Gly216 carbonyl in the model of the wild-type sequence (Fig. 5C), whereas in the mutant model this H-bond is disrupted by slight rotation of Val198 to optimize hydrophobic contacts with the Pro17 ring and with the β-carbon of Arg217A (Fig. 5D). The impact of the mutation on binding interactions with CTRC suggested by these models would not appear to be sufficient to explain the functional importance of the mutation in predisposing carriers to pancreatitis.

Consistent with these minimal structural differences, the calculated binding energy for the p.A16V mutant sequence was very close to that of the wild-type sequence (−8.811 versus −10.16 kcal/mol, respectively) and did not suggest enhanced binding upon mutation as might be predicted by functional studies (4, 5). One possible explanation for this apparent discrepancy is that the energy calculations do not take into account entropic components of binding energy, which may differ between the substrate sequences. This explanation is highly plausible, because entropic factors involved in binding are anticipated to differ between the wild-type and mutant activation peptides as detailed further under “Discussion.” Thus, it is probable that the mutation critically influences the conformational ensemble and dynamics of the activation peptide in its unbound form, resulting in differential entropic contributions to formation of the CTRC-substrate complex and thereby altering cleavage selectivity.

**DISCUSSION**

On one hand, biochemical studies with purified proteins show that CTRC, like most digestive proteases, is capable of proteolytic cleavage of multiple target sequences revealing no highly conserved recognition motif (Table 1). CTRC appears to be relatively insensitive to a variety of mutations of the trypsinogen Ca2+-binding loop cleavage site (16), further suggesting fairly promiscuous activity. On the other hand, incontrovertible genetic evidence demonstrates that a subtle mutation of the CTRC cleavage site within the cationic trypsinogen activation peptide predisposes to development of chronic pancreatitis (7, 8), as does loss of function of CTRC itself (9–12); these observations reveal a specific regulatory role for the enzyme.

To reconcile these contrasting views of CTRC specificity, we consider how the protease selects its substrates in vivo. Steady state enzyme kinetics studies using individual substrates can reveal the thermodynamic stability of the Michaelis complex ($K_m$) and the overall catalytic rate of the reaction ($k_{cat}$) from which the specificity constant $k_{cat}/K_m$ is derived. However, in the scenario in which multiple candidate substrates are in direct competition, kinetic rather than thermodynamic control may prevail (57, 58). Protein-protein association is a multistep process, with different molecular forces influencing the rates of each step, and electrostatic interactions provide the dominant long-range force capable of accelerating the rate of initial molecular collision (58, 59). The intense concentration of positive charge surrounding the CTRC active site, yet apparently not positioned for optimal formation of tight salt bridges in the final Michaelis complex, suggests that this funnel-like ring of surface charge may instead be optimized to kinetically favor the initial attraction of polyacidic substrate sequences, akin to a molecular tractor beam.

Non-specific long-range electrostatic attraction will be a dominant driving force as CTRC and its substrate first approach each other, at a stage when they may not yet be correctly rotationally oriented. The stabilization conferred by electrostatic attraction, combined with the viscosity of the surrounding solution, may also extend the lifespan of the low affinity transient complex, allowing the two proteins to rotate and sample multiple trajectories of approach in repeated microcollisions (58). Thus, an electrostatically complementary substrate will have enhanced probability of finding a productive binding conformation, achieving the Michaelis complex, and becoming proteolyzed. A similar influence of electrostatic potential on substrate specificity was observed previously for the hepatitis C virus NS3 protease using pre-steady state kinetics, where clusters of positively charged residues near the active site, complementary to clusters of negative charge on the substrate, were found to drive very rapid association (60). The influence of these long-range interactions on substrate selection in vivo may be underestimated in comparisons of overall substrate affinity, which depend upon the relative rates of formation and dissociation of the high affinity Michaelis complex. The later stages of binding that mark the progression from transient to Michaelis complex are likely to be slower, due to the radical alterations of local substrate conformation often required (56).

Following formation of a roughly aligned encounter complex, the next stage of CTRC-substrate association is likely to be docking of the large hydrophobic P1 “anchor” residue within the S1 subsite (57). As with electrostatic charge, the kinetic importance of this interaction and its impact on specificity under conditions of direct substrate competition may be underestimated in studies with single purified substrates. For example, CTRC binding studies with tight binding inhibitors revealed a clear preference for P1 Leu over Met, both of which were represented much more highly than Phe or Tyr in a pool of inhibitors selected by phage display (17). By contrast, CTRC appeared insensitive to mutation of the P1 Leu in the trypsinogen Ca2+-binding loop to Met, Tyr, or Phe (16). We would speculate that in the case of the rigidly structured inhibitors, the overall rate of association reflects the rate of $P_1$-$S_1$ docking, which is fastest for Leu due to optimal complementarity with the $S_1$ pocket. By contrast, with highly flexible or natively unstructured substrate sequences, the slower rate with which flanking residues conform to adjacent subsites through induced fit (57) may mask the importance of $P_1$ specificity.

CTRC serves a protective function in the pancreas, as demonstrated by the association of CTRC loss of function with chronic pancreatitis (9–12), and by the finding that pathological activation of hereditary pancreatitis-causing cationic trypsinogen mutants is dependent on CTRC activity (4). We have found that CTRC can cleave cationic trypsinogen at alternative cleavage sites; proteolysis within the trypsinogen activation peptide leads to enhanced autoactivation of trypsin (5), whereas...
proteolysis within the $\text{Ca}^{2+}$-binding loop leads to trypsinogen degradation (6). The protective function of CTRC would suggest that the dominant activity of CTRC activated in the pancreas must be the degradation of cationic trypsinogen, initiated by cleavage within the $\text{Ca}^{2+}$-binding loop.

Our modeling results are consistent with the idea that when sterically accessible, the $\text{Ca}^{2+}$-binding loop is the more preferred site of cleavage. This site fills a greater number of nonprimed subsites, burying a greater solvent accessible surface area, it possesses the preferred Leu at the anchor $P_1$ position, and it is calculated to have a thermodynamically more favorable binding energy. In the absence of bound $\text{Ca}^{2+}$, the $\text{Ca}^{2+}$-binding loop may be preferred by CTRC from the perspective of long-range electrostatic interactions as well, because it possesses four Glu residues within the primary sequence of the loop and an additional three exposed acidic residues on nearby loops. In comparison, the unstructured activation peptide possesses four tandem Asp residues, but the negative charge of these side chains would be offset by an adjacent Lys residue and by the $N$-terminal amine.

It is more difficult to explain from a structural perspective why the p.A16V mutant trypsinogen activation peptide is improved as a CTRC substrate, to the extent of significantly diminishing the protective function of CTRC in the pancreas. However, it may be as a result of the impact of this substitution on the conformational ensemble represented by the activation peptide and on its dynamics. The activation peptide is at the N terminus of trypsinogen, and is unstructured in the crystal structures of bovine and rat trypsinogens (61–63). In this unstructured state, the Ala$^{16}$-Pro$^{17}$ (or Val$^{16}$-Pro$^{17}$) peptide bond exists as an equilibrium mixture of cis and trans isomers (64, 65), and the bulkier Val$^{16}$ will greatly reduce the rate of cis-trans isomerization (66). Perhaps more importantly, the steric bulk of the proline ring restricts the conformations available to the preceding residue (67), and the $\beta$-branched Val$^{16}$ will encounter further restrictions than the smaller Ala$^{16}$ (68). To bind to the CTRC active site and undergo proteolysis, the Pro peptide bond must assume the trans configuration (69), and while both residues must conform to the idealized binding mode of a canonical loop, in which the $P_1$ residue main chain assumes an antiparallel $\beta$-strand conformation and the $P_2$ residue main chain assumes a polyproline II conformation (43). We speculate that the steric restrictions and reduced mobility of the p.A16V mutant have the effect of locking the activation peptide into a more substrate-like conformation, or of biasing the conformational ensemble toward a more substrate-like conformation. The impact would be to minimize the unfavorable entropic contribution to the binding energy, rendering the p.A16V mutant trypsinogen activation peptide a more favorable substrate.

In conclusion, the molecular structure and analyses of CTRC presented here offer mechanistic insight into the striking selectivity of the enzyme for regulatory sites in trypsinogens and procarboxypeptidases. Long-range electrostatic interactions between enzyme and substrate, rather than specific charge pairing, underlie substrate discrimination. In addition, the structure of CTRC establishes a framework that may enable interpretation of the functional effects of other clinically significant trypsinogen mutations that impact CTRC substrate selectivity, as well as mutations within CTRC itself that modify risk for development of chronic pancreatitis.

Acknowledgments—Diffraction data were measured at beamline X29 of the National Synchrotron Light Source, which is supported by the Offices of Biological and Environmental Research and Basic Energy Sciences of the United States Department of Energy, and the National Center for Research Resources of the National Institutes of Health.

REFERENCES

1. Rinderknecht, H. (1986) Activation of pancreatic zymogens. Normal activation, premature intrapancreatic activation, protective mechanisms against inappropriate activation. *Dig. Dis. Sci.* **31**, 314–321.
2. Whitcomb, D. C., Gorry, M. C., Preston, R. A., Furey, W., Sossenheimer, M. J., Ulrich, C. D., Martin, S. P., Gates, L. K., Jr., Amann, S. T., Toskes, P. P., Liddle, R., McGrath, K., Uomo, G., Post, J. C., and Ehrlich, G. D. (1996) Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat. Genet.* **14**, 141–145.
3. Teich, N., Rosendahl, J., Töth, M., Mössner, J., and Sahin-Töth, M. (2006) Mutations of human cationic trypsinogen (PRSS1) and chronic pancreatitis. *Hum. Mutat.* **27**, 721–730.
4. Szabó, A., and Sahin-Töth, M. (2012) Increased activation of hereditary pancreatitis-associated human cationic trypsinogen mutants in presence of chymotrypsin C. *J. Biol. Chem.* **287**, 20701–20710.
5. Nemoda, Z., and Sahin-Töth, M. (2006) Chymotrypsin C (caldecin) stimulates autocatalytic activation of human cationic trypsinogen. *J. Biol. Chem.* **281**, 11879–11886.
6. Szmola, R., and Sahin-Töth, M. (2007) Chymotrypsin C (caldecin) promotes degradation of human cationic trypsin. Identity with Rinderknecht’s enzyme Y. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11227–11232.
7. Witt, H., Luck, W., and Becker, M. (1999) A signal peptide cleavage site mutation in the cationic trypsinogen gene is strongly associated with chronic pancreatitis. *Gastroenterology* **117**, 7–10.
8. Grocock, C. J., Rebourcs, V., Delhaye, M. N., Andrien-Sandberg, A., Weiss, F. U., Mountford, R., Harcus, M. J., Niemczyk, E., Vitone, L. I., Dodd, S., Jørgensen, T. M., Ammann, R. W., Schaffalitzky de Muckadell, O., Butler, J. V., Burgess, P., Kerr, B., Charnley, R., Sutton, R., Ratary, M. G., Devière, J., Whitcomb, D. C., Neoptolemos, J. P., Lévy, P., Lerch, M. M., Greenhalgh, W., European Registry of Hereditary Pancreatitis and Pancreatic Cancer (2010) The variable phenotype of the p.A16V mutation of cationic trypsinogen (PRSS1) in pancreatitis families. *Gut* **59**, 357–363.
9. Rosendahl, J., Witt, H., Szmola, R., Bhata, E., Özsvari, B., Landt, O., Schulz, H. U., Gress, T. M., Pützer, R., Lühr, M., Kovacs, P., Blüher, M., Stumvoll, M., Choudhuri, G., Hegyi, P., te Morsche, R. H., Drenth, J. P., Truninger, K., Macek, M., Jr., Puhl, G., Witt, U., Schmidt, H., Bünning, C., Ockenga, J., Kage, A., Groneberg, D. A., Nickel, R., Berg, T., Wiedemann, B., Bödeker, H., Keim, V., Mössner, J., Teich, N., and Sahin-Töth, M. (2008) Chymotrypsin C (CTRC) variants that diminish activity or secretion are associated with chronic pancreatitis. *Nat. Genet.* **40**, 78–82.
10. Masson, E., Chen, J. M., Scotet, V., Le Maréchal, C., and Féric, C. (2008) Association of rare chymotrypsinogen C (CTRC) gene variations in patients with idiopathic chronic pancreatitis. *Hum. Mutat.* **31**, 83–91.
11. Zhou, J., and Sahin-Töth, M. (2011) Chymotrypsin C mutations in chronic pancreatitis. *J. Gastroenterol. Hepatol.* **26**, 1238–1246.
12. Beer, S., Zhou, J., Szabó, A., Keiles, S., Chandak, G. R., Witt, H., and Sahin-Töth, M. (September 1, 2012) Comprehensive functional analysis of chymotrypsin C (CTRC) variants reveals distinct loss-of-function mechanisms associated with pancreatitis risk. *Gut* **10.1136/gutjnl-2012–303090*.
13. Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501–4524.
14. Perona, J. I., and Craik, C. S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337–360.
60. Koch, U., Biasiol, G., Brunetti, M., Fattori, D., Pallaoro, M., and Steinkühler, C. (2001) Role of charged residues in the catalytic mechanism of hepatitis C virus NS3 protease. Electrostatic precollision guidance and transition-state stabilization. *Biochemistry* **40**, 631–640

61. Pasternak, A., Ringe, D., and Hedstrom, L. (1999) Comparison of anionic and cationic trypsinogens. The anionic activation domain is more flexible in solution and differs in its mode of BPTI binding in the crystal structure. *Protein Sci.* **8**, 253–258

62. Fehlhammer, H., Bode, W., and Huber, R. (1977) Crystal structure of bovine trypsinogen at 1–8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J. Mol. Biol.* **111**, 415–438

63. Kossiakoff, A. A., Chambers, J. L., Kay, L. M., and Stroud, R. M. (1977) Structure of bovine trypsinogen at 1.9-Å resolution. *Biochemistry* **16**, 654–664

64. Grathwohl, C., and Wüthrich, K. (1981) NMR studies of the rates of proline cis–trans isomerization in oligopeptides. *Biopolymers* **20**, 2623–2633

65. Wedemeyer, W. J., Welker, E., and Scheraga, H. A. (2002) Proline cis-trans isomerization and protein folding. *Biochemistry* **41**, 14637–14644

66. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**, 4953–4963

67. MacArthur, M. W., and Thornton, J. M. (1991) Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**, 397–412

68. Schimmel, P. R., and Flory, P. J. (1968) Conformational energies and configurational statistics of copolypeptides containing L-proline. *J. Mol. Biol.* **34**, 105–120

69. Fischer, G. (1994) Peptidyl-prolyl cis/trans isomerases and their effectors. *Angew. Chem. Int. Ed. Engl.* **33**, 1415–1436

70. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162