THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 14, pp. 9848–9859, April 5, 2013
Published in the U.S.A.

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

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], András 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.

Human chymotrypsin C (CTRC) is a pancreatic serine protease that regulates activation and degradation of trypsinogens and procarboxypeptidases by targeting specific cleavage sites within their zymogen precursors. In cleaving these regulatory sites, which are characterized by multiple flanking acidic residues, CTRC shows substrate specificity that is distinct from that of other isoforms of chymotrypsin and elastase. Here, we report the first crystal structure of active CTRC, determined at 1.9-Å resolution, revealing the structural basis for binding specificity. The structure shows human CTRC bound to the small protein protease inhibitor eglin c, which binds in a substrate-like manner filling the S_6-S_5' subsites of the substrate binding cleft. Significant binding affinity derives from burial of preferred hydrophobic residues at the P_1, P_4, and P_2' positions of CTRC, although acidic P_2' residues can also be accommodated by formation of an interfacial salt bridge. Acidic residues may also be specifically accommodated in the P_6 position. The most unique structural feature of CTRC is a ring of intense positive electrostatic surface potential surrounding the primarily hydrophobic substrate binding site. Our results indicate that long-range electrostatic attraction toward substrates of concentrated negative charge governs substrate discrimination, which explains CTRC selectivity in regulating active digestive enzyme levels.

Digestive proteases are synthesized and secreted by the pancreas as inactive zymogens. Physiological activation takes place in the duodenum, where enteropeptidase initiates an activation cascade by specifically activating trypsinogens, which in turn activate chymotrypsinogens, proelastases, procarboxypeptidases, and other digestive enzymes (1). Premature activation of trypsin within the pancreas is understood to be a major initiating factor in chronic pancreatitis, and chymotrypsin C (CTRC)[4] is a significant player in this process. In approximately half of the families affected by autosomal dominant hereditary pancreatitis, the disease is caused by mutations in the cationic trypsinogen gene *PRSS1* that result in either enhanced trypsinogen activation or resistance to degradation (2–4). CTRC possesses the unique capacity to impact trypsinogen activation and stability via two opposing mechanisms: it can cleave cationic trypsinogen either at Phe^{18}-Asp^{19} within the trypsinogen activation peptide, leading to enhanced autoactivation (5), or at Leu^{81}-Glu^{82} within the Ca^{2+}-binding loop, resulting in degradation (6). A number of disease-causing cationic trypsinogen mutations exert their effect in part through accelerating cleavage by CTRC at Phe^{18}-Asp^{19} or through diminishing cleavage by CTRC at Leu^{81}-Glu^{82} (4). The p.A16V mutation, which accounts for a small percentage of hereditary pancreatitis kindreds and is also associated with idiopathic chronic pancreatitis (7, 8), appears to exert its pathological effect solely by increasing the vulnerability of the cationic trypsinogen activation peptide to cleavage by CTRC (4). Finally, several mutations in the *CTRC* gene itself that lead to loss or impairment of protein function are significantly associated with chronic pancreatitis (9–12).

Both cationic trypsin and CTRC are members of the chymotrypsin family of serine peptidases, which share a common two β-barrel-fold, a famous triad of catalytic residues Ser, His, and Asp, and a conserved catalytic mechanism for nucleophilic cleavage of peptide bonds. Although these enzymes all catalyze the same reaction, they are differentiated by the distinct substrate sequences that they recognize through a series of subsites

* This work was supported, in whole or in part, by National Institutes of Health Grants R01 DK082412-S2 (ARRA), R01 DK082412, R01 DK058088, and R01 DK095753 (to M. S.-T.) and Florida Department of Health Postdoctoral Fellowship 1BD-01 (to J. B.).
[S] This article contains supplemental Models S1–S3.
[1] Both authors contributed equally to this work.
[2] To whom correspondence may be addressed: 72 East Concord St., Evans-433, Boston, MA 02118. E-mail: miklos@bu.edu.
[3] To whom correspondence may be addressed: 310 Griffin Bldg., 4500 San Pablo Rd., Jacksonville, FL 32224. E-mail: radisky.evette@mayo.edu.

[4] The abbreviations used are: CTRC, chymotrypsin C; PDB, Protein Data Bank.

TABLE 1
Specific substrate/inhibitor sequences recognized by CTRC

Substrate cleavage site	P_6	P_5	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'	P_5'	ref
*Tg1, Tg2 activation peptide (F^{18}-D^{19})	-	-	-	A	P	F	**D**	**D**	**D**	**D**	K	(5)
*Tg1 p.A16V activation peptide (F^{18}-D^{19})	-	-	-	V	P	F	**D**	**D**	**D**	**D**	K	(5)
proCPA1 activation peptide (L^{96}-L^{97})	**E**	**D**	V	Q	S	L	L	**D**	**E**	**E**	Q	(15)
proCPA2 activation peptide (L^{96}-L^{97})	**E**	**D**	V	Q	S	L	L	**D**	K	**E**	N	(15)
*Tg1, Tg2 Ca^{++} binding loop (L^{81}-E^{82})	H	N	I	**E**	V	L	**E**	G	N	**E**	Q	(6)
Tg3 Ca^{++} binding loop (L^{81}-E^{82})	H	N	I	K	V	L	**E**	G	N	**E**	Q	(6)
Tg1, Tg2 alternative site (L^{41}-N^{42})	P	Y	Q	V	S	L	N	S	G	Y	H	(6)
Tg2 alternative site (Y^{37}-Q^{38})	**E**	N	S	V	P	Y	Q	V	S	L	N	(6)
Tg2 alternative site (L^{189}-E^{190})	F	C	V	G	F	L	**E**	G	G	K	**D**	(6)
Tg3 alternative site (L^{148}-S^{149})	G	W	G	N	T	L	S	F	G	A	**D**	(6)
Tg3 alternative site (Y^{150}-G^{151})	G	N	T	L	S	F	G	A	**D**	Y	P	(6)
Inhibitor reactive site	**P_6**	**P_5**	**P_4**	**P_3**	**P_2**	**P_1**	**P_1'**	**P_2'**	**P_3'**	**P_4'**	**P_5'**	**ref**
eglin c (L^{45}-D^{46})	G	S	P	V	T	L	**D**	L	R	Y	N	-
SGPI-2 (L^{30}-K^{31})	S	A	A	C	T	L	K	A	C	P	Q	(17)
**SGPI-2 library selection consensus	(S)	(A)	A/G…	(C)	T/S	M/L/F/Y	M/L/A/K	L/D/E…	(C)	**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 P_4, P_2, P_1, P_1', P_2', and P_4'. 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 P_1 residues,[5] and ~40% sequence identity with other human chymotrypsin isoforms, which typically cleave after aromatic P_1 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 Ca^{2+}-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 cationic trypsinogen, with somewhat contradictory results (16, 17). Inhibitor studies suggest strong preference for Leu followed by Met at the P_1 position (17), whereas substrate mutagenesis suggests that CTRC is much more permissive of alternative P_1 residues (16). Another potential specificity feature is suggested by the presence of multiple Asp and Glu residues 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 confirms a strong preference for acidic residues at the P_4' position (17), but mutagenesis studies again show only moderate effects upon alteration of individual charged residues within the cationic trypsinogen Ca^{2+}-binding loop (16).

Here, the first structure of active CTRC reveals a familiar fold with distinctive electrostatic features surrounding the substrate binding 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 P_1 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 P_1 position, and may also be influenced by local sequence-dependent backbone conformational tendencies.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—Human chymotrypsinogen C bearing a C-terminal His_{10} 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-

[5] Substrate residues surrounding the cleavage site are designated by the nomenclature of Schechter and Berger (70): starting from the scissile bond, substrate residues are numbered P_1, P_2, P_3, etc. in the direction of the N terminus (collectively the non-primed residues), and P_1', P_2', P_3', etc. in the direction of the C terminus (collectively the primed residues). Corresponding enzyme subsites are numbered S_1, S_2, S_3, etc.

harose chromatography essentially as previously described (18). Eglin c was dialyzed into 10 mM $(NH_4)_2OAc$ at pH 6.0, lyophilized, and stored at −80 °C until use, when it was reconstituted with H_2O.

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 × 0.2 × 0.2 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 N_2.

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 $P2_12_12_1$, 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 1ACB, 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 ($F_o - F_c$) 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 ($2F_o - F_c$) 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 His_{10} tag or the *N*-linked glycosyl groups of CTRC due to the absence of electron density. The final R/R_{free} was 0.157/0.210. Figs. 1, 2, 4*A*, 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 (CTRB1, CTRB2, 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 (30) 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	$P2_12_12_1$
Unit cell, Å	56.27, 76.25, 81.82 90°, 90°, 90°
Resolution, Å	1.9
Unique reflections	27,661
Completeness (%)	97.34 (76.8)[a]
Multiplicity	6.8 (4.8)[a]
Mean I/ (σ)	51.3 (9.0)[a]
R_{merge}	0.043 (0.146)[a]
R_{sym}	0.043 (0.146)[a]
$R_{\text{cryst}}/R_{\text{free}}$ (%)	0.157/0.210
Root mean square deviation bonds, Å	0.027
Root mean square deviation angles (degree)	2.066
Protein atoms	2469
Ions	3 (PO_4^{-2})
Water molecules	320
Φ,Ψ angle distribution[b]	
In favored regions	209 (87.8%)
In additionally allowed regions	28 (11.8%)
In generously allowed regions	1 (0.4%)

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

scoring features, was selected for visual and energetic comparison (31, 33, 34).

RESULTS

Structure Solution and Refinement—The structure of active human CTRC bound to inhibitor eglin c was determined to 1.9-Å resolution in space group $P2_12_12_1$ 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 search model as a result of substantial conformational changes that take place upon protease activation (36). The model was rebuilt and refined to a final R_{cryst} (R_{free}) 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 Ser^{195} (Ser^{216}),[6] His^{57} (His^{74}), and Asp^{102} (Asp^{121}) (Fig. 1*A*). The substrate binding cleft between the β-barrels is occupied by bound inhibitor eglin c, a 70-amino acid protein 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. 1*B*).

CTRC residues 1–10 of the cleaved activation peptide chain (CGVPSFPPNL) are retained by the activated enzyme due to a disulfide link between Cys^1 (Cys^{17}) of the activation peptide and Cys^{122} (Cys^{141}), located in the linker between the two β-barrels on the enzyme face distal from the active site (Fig. 1*C*). 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 Gly^2 (Gly^{18}), an H-bond between the carbonyl of Pro^7 (Pro^{23}) and the side chain of Arg^{23} (Arg^{37}), and substantial hydrophobic interactions of Pro^4 (Pro^{20}), Phe^6 (Phe^{22}), Pro^8 (Pro^{24}), and Leu^{10} (Leu^{26}) (Fig. 1*C*). Clear density is observed for the C-terminal Leu^{10} (Leu^{26}) of the activation peptide chain, revealing that residues Ser^{11}-Ala^{12}-Arg^{15} (Ser^{27}-Ala^{28}-Arg^{29}) of the activation peptide have been proteolytically removed. Given the binding preference of CTRC for Leu at the P_1 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 His_{10} 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 Asn^{36B} (Asn^{52}), a modification required for efficient folding and secretion (48), we did not observe density for the glycosyl group. The side chain of Asn^{36B} (Asn^{52}) was poorly defined and it is likely that the sugar 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 Arg^{15}-Val^{16} (Arg^{29}-Val^{30}), the new N-terminal residue Val^{16} (Val^{30}) becomes buried, forming a salt bridge with Asp^{194} (Asp^{215}) and an H-bond with the carbonyl oxygen of Arg^{143} (Arg^{162}). 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 S_1 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 chymotrypsinogen C, small conformational adjustments in all three members of the triad bring them into appropriate alignment for catalysis, reducing the Ser^{195} Oγ-His^{57} Nϵ2 distance from 3.83 to 2.88 Å, and shortening the His^{57} Nδ1-Asp^{102} Oδ2 H-bond from 2.95 to 2.55 Å. These alterations are typical of the conserved mechanism of activation of chymotrypsinogen family members (13).

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

[6] The CTRC residue numbering used in this report and in the crystal structure coordinates is derived by homology to bovine chymotrypsin, the archetypal 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.

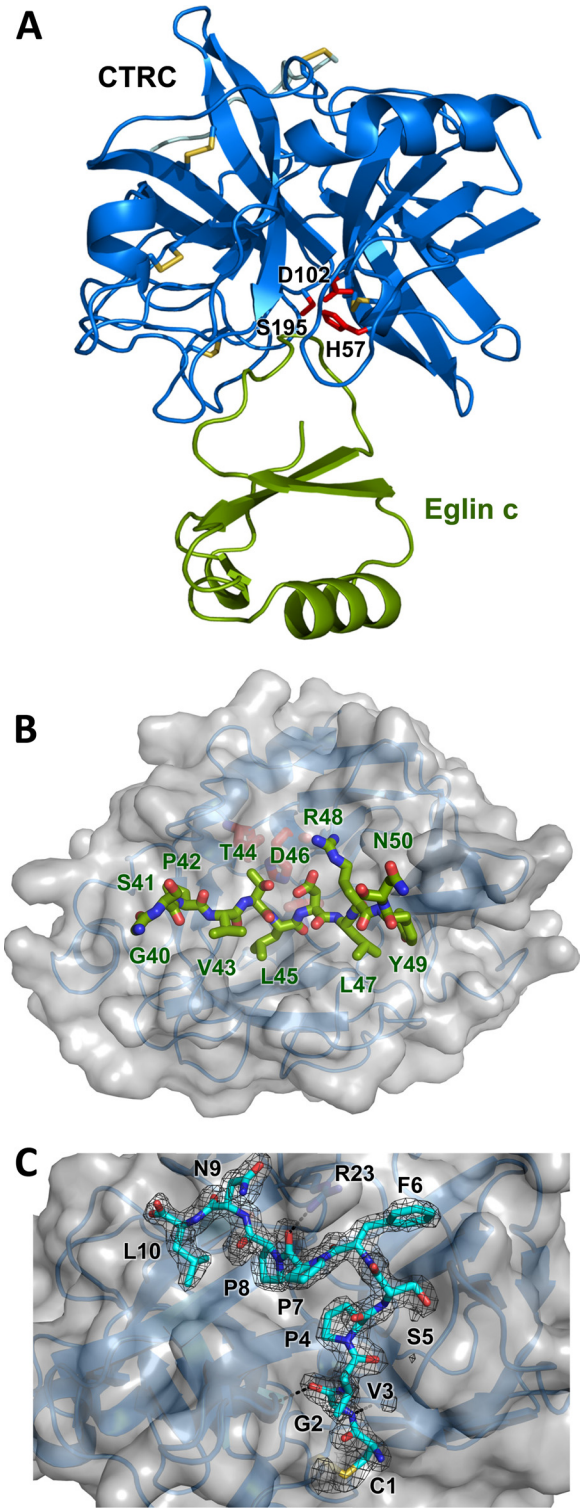

FIGURE 1. **Crystal structure of the CTRC-eglin c complex.** *A*, structural overview of the complex. CTRC is shown in *blue*, with catalytic triad residues Ser^{195}, His^{57}, and Asp^{102} in *red* and disulfide links in *yellow*. Eglin c is displayed in *green*. *B*, view into the substrate binding cleft of CTRC. CTRC is shown with a *semitransparent gray surface*. Eglin c binding loop residues 40–50 are rendered in stick representation, filling (from left to right) CTRC S_6-S_1 and S_1'-S_5' subsites. *C*, retained activation peptide of CTRC. Residues 1–10 of the chymotrypsin C activation peptide are tethered to the activated enzyme through a disulfide link between Cys^1 and Cys^{122}. The activation peptide is depicted in *cyan* in stick representation, with a $2F_o - F_c$ electron density map contoured at 1.6σ. Strong density around the Leu^{10} carboxyl terminus confirms that residues 11–13 of the activation peptide are not disordered but have been proteolytically removed.

a hydrophobic "mini-core" (49) comprised of Leu^{37}, Val^{43}, and Phe^{55}, and by a H-bond network involving Thr^{44}, Asp^{46}, the Arg^{48} backbone N, Arg^{51}, Arg^{53}, and the C-terminal carboxyl group of Gly^{70} (Fig. 2*A*). 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 Leu^{45}-Asp^{46} reactive site peptide bond of eglin c is appropriately positioned for nucleophilic attack by the catalytic Ser^{195} (Ser^{216}) of CTRC (Fig. 2*B*), 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 P_1-P_6 and primed side residues P_1'-P_5'. The non-primed side residues are positioned by antiparallel backbone-backbone H-bonds between the P_3 residue Val^{43} and CTRC Gly^{216} (Gly^{238}), and between the P_1 residue Leu^{45} and CTRC Ser^{214} (Ser^{236}) (Fig. 2*C*). The Leu^{45} carbonyl oxygen is positioned to interact with the CTRC oxyanion hole amide nitrogens of Ser^{195} (Ser^{216}) and Gly^{193} (Gly^{214}). On the primed side, additional H-bonds are formed between P_2' residue Leu^{47} and CTRC Thr^{41} (Thr^{58}) (Fig. 2*C*).

The S_1 primary specificity pocket of CTRC, occupied in the complex by the side chain of eglin c Leu^{45}, is shaped by the hydrophobic side chains of CTRC residues Ala^{190} (Ala^{211}), Val^{213} (Val^{235}), and Val^{226} (Val^{250}) (Fig. 2*C*). A more shallow hydrophobic depression in the substrate binding cleft is shaped by the side chains of CTRC Leu^{99} (Leu^{118}) and Phe^{215} (Phe^{237}), which form a binding pocket for P_4 residue Pro^{42}. On the primed side, the S_2' subsite, a pocket bordered by the basic side chain of CTRC Arg^{143} (Arg^{162}) and the hydrophobic side chain of Ile^{151} (Ile^{169}), is filled by the hydrophobic P_2' residue Leu^{47} (Fig. 2*C*).

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 Arg^{175} (Arg^{195}), Arg^{218} (Arg^{241}), and Lys^{224} (Lys^{248}) (Fig. 2*D*). This basic pocket forms the S_6 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 P_6 residue Gly^{40} (Fig. 2*D*).

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 when compared with other chymotrypsin and elastase isoforms (16). By contrast, the elastase isoforms show broad P_1 specificity but comparatively low catalytic efficiency on short peptide substrates, whereas chymotrypsins prefer aromatic residues Phe, Tyr, or Trp at P_1 (16). The position occupied in CTRC by Val^{226} (Val^{250}), one of the hydrophobic residues shaping the S_1 pocket (Fig. 2*C*), is in other chymotrypsins filled by Gly or Ala, and the

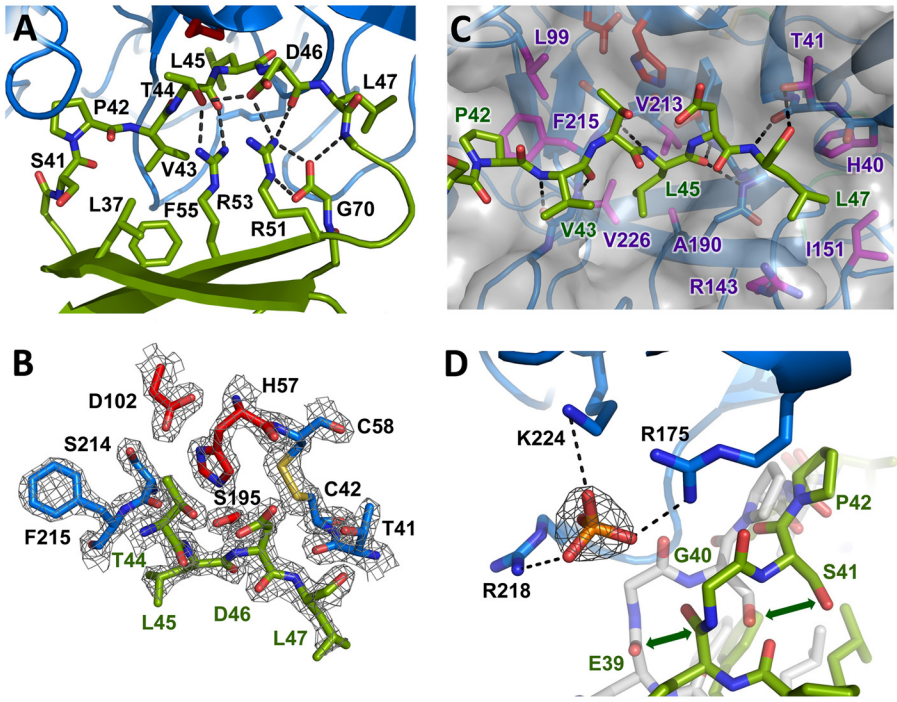

FIGURE 2. **Eglin c inhibitory interaction with CTRC.** *A*, stabilization of the inhibitory loop of eglin c. The eglin c binding loop assumes a substrate-like canonical conformation of the peptide backbone, stabilized on the nonprimed side by hydrophobic interactions of Leu[37], Val[43], and Phe[55], and on the primed side by an H-bond network involving Thr[44], Asp[46], the Arg[48] amide nitrogen, Arg[51], Arg[53], and C-terminal Gly[70]. The side chain of Arg[48] is omitted for clarity. *B*, the CTRC-eglin c complex resembles an enzyme-substrate Michaelis complex. The Leu[45]-Asp[46] reactive site peptide bond of eglin c, linking the P_1 and P_1' residues, lies in proper orientation for attack by the catalytic Ser[195] of CTRC. The $2F_o - F_c$ electron density map is shown contoured at 2.0σ. *C*, key CTRC-eglin c binding interactions. The eglin c P_1 residue Leu[45] fills the S_1 pocket bordered by CTRC Ala[190], Val[213], and Val[226]. P_4 residue Pro[42] fills a hydrophobic concavity formed by CTRC Leu[99] and Phe[215]. P_2' residue Leu[47] fills a pocket formed by CTRC Arg[143] and Ile[151]. Multiple backbone H-bonds orient the inhibitor, indicated by *black dotted lines*. *D*, positively charged P_6 pocket displaces eglin c backbone to bind phosphate. Basic side chains of CTRC Arg[175], Arg[218], and Lys[224] 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 $2F_o - F_c$ electron density map shown for the phosphate ion is contoured at 1.6σ.

bulkier Val residue at this position is likely to be responsible for the modest binding selectivity of CTRC for Leu in preference to Met, Phe, or Tyr at the P_1 position, as identified by phage display selection and inhibitor binding studies (17) and by K_m values for cleavage of tetrapeptide substrates (16). Nevertheless, the S_1 subsite of CTRC is still capable of accommodating aromatic residues, and in fact, CTRC catalytic rates are slightly enhanced for cleavage after these bulkier residues, resulting in comparable catalytic efficiencies for cleavage after Leu, Met, Phe, or Tyr (16). This result is also consistent with the identification of natural CTRC cleavage sites within protein substrates after Phe, Leu, and Tyr (Table 1).

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 P_4' position, an element of specificity corroborated by phage display selection (17). Acidic residues can also be found at P_1', P_2', P_3', and P_5' on the primed side of the cleavage site, and at P_3, P_5, and P_6 on the non-primed side of the cleavage site (Table 1). To gain insight into the probable electrostatic contribution to this unusual substrate specificity, we calculated the predicted electrostatic surface potential of the CTRC structure, and for comparison we generated homology models and calculated electrostatic surfaces for other human chymotrypsin and elastase isoforms, which do not target the same regulatory cleavage sites (Fig. 3). We observed a particularly striking concentration of positive charge in a ring surrounding the substrate binding cleft of CTRC (Fig. 3, *top*). 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 P_5 and P_6 substrate residues, and a second cluster of basic residues on the primed side bordering the subsites that recognize P_2', P_3', and P_4' substrate residues. Notably, this charge distribution contrasts markedly with the predicted electrostatic surfaces of other elastase and chymotrypsin isoforms (Fig. 3, *lower panels*). The human elastases feature small patches of both positive and negative charge in roughly equal proportions, whereas human chymotrypsins 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.

Modeling of CTRC-Substrate Complexes—Selection by CTRC among several potential cleavage sites within human cationic trypsinogen has profound health implications. Whereas cleavage in the trypsinogen Ca^{2+}-binding loop is a normal mechanism for CTRC to apply the brakes to a premature cascade of digestive enzyme activation in the pancreas (6), an alternative cleavage within the activation peptide has the potential to accelerate this activation cascade (5). To serve its protective function, CTRC must significantly favor the former cleavage over the latter in the pancreas environment. To gain insight into the interactions of competing cleavage sites with CTRC, we examined the structural context of the Ca^{2+}-binding loop

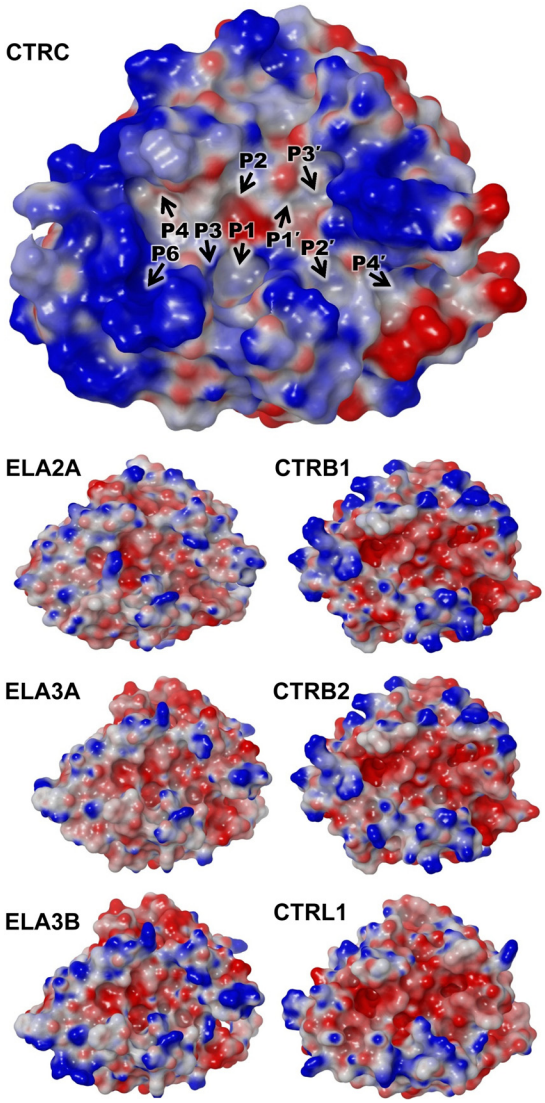

FIGURE 3. **Distinct highly charged electrostatic surface of CTRC.** The electrostatic surface potential of CTRC (*top*) shows intense concentration of a positive charge (*blue*) surrounding the active site cleft. Positions of substrate-binding subsites are indicated by *black arrows*. *Smaller panels below* show for comparison the electrostatic surface potentials generated for homology models of human elastase isoforms (*left*) and chymotrypsin isoforms (*right*). Molecular surfaces were generated using the Molecular Surface module of Schrodinger 2012 as described under "Experimental Procedures"; the electrostatic potential color ramp was set from a minimum of −0.35 to a maximum of 0.15.

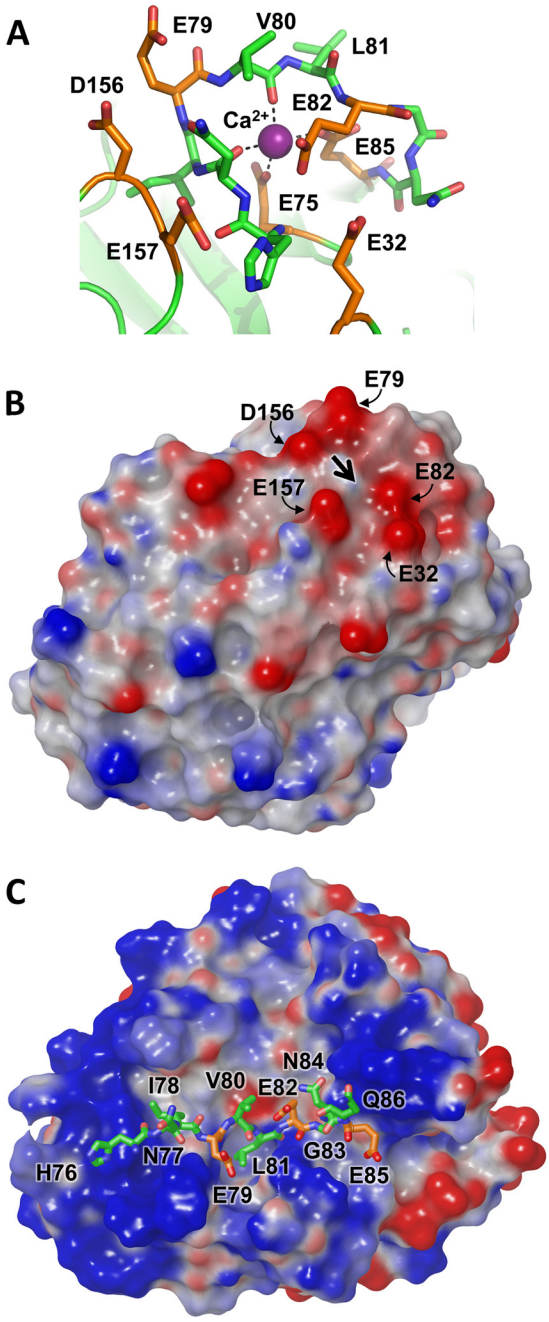

FIGURE 4. **Cationic trypsinogen Ca^{2+}-binding loop targeted for cleavage by CTRC.** *A*, the Ca^{2+}-binding loop of human cationic trypsinogen contains multiple acidic residues (*orange*), with additional acidic residues located in nearby neighboring loops. The bound Ca^{2+} is shown in *purple*; the Leu^{81}-Glu^{82} peptide bond represents the preferred cleavage site targeted by CTRC. Structure coordinates are from PDB code 2RA3 (55). *B*, the electrostatic surface potential calculated for cationic trypsin reveals a negative charge surrounding the site of bound Ca^{2+}, indicated by the *large black arrow*. Molecular surfaces were generated using the Molecular Surface module of Schrodinger 2012 as described under "Experimental Procedures"; the electrostatic potential color ramp was set from a minimum of −0.35 to a maximum of 0.15. *C*, Ca^{2+}-binding loop cleavage sequence of cationic trypsinogen is shown modeled into the active site cleft of CTRC. The hydrophobic side chains of Ile^{78} and Leu^{81} occupy the complementary S_4 and S_1 subsites, respectively. Acidic residues are shown in *orange*; Glu^{79} at the P_2 position and Glu^{85} at the P_4' position are near CTRC regions of positive surface potential.

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^{2+} is coordinated within the loop by the side chains of residues Glu^{75} and Glu^{85}, which anchor the base of the loop, and by carbonyl oxygens of Asn^{77} and Val^{80} (Fig. 4*A*).[7] CTRC targets the Leu^{81}-Glu^{82} peptide bond for cleavage. With Ca^{2+} bound, Leu^{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

[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^{75} and Glu^{85}, there are two additional glutamates in the loop (Glu^{79} and Glu^{82}), as well as three acidic residues in neighboring loops (Glu^{32}, Asp^{156}, and Glu^{157}; Fig. 4*A*). Thus, even with Ca^{2+} present to neutralize the charge of Glu^{75} and Glu^{85}, there is a strongly negative electrostatic potential covering this region of the molecule (Fig. 4*B*). This potential will generate macroscopic electrostatic complementarity between the Ca^{2+}-binding loop of cationic trypsin or trypsinogen and the substrate-binding site of CTRC. It is anticipated that when Ca^{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^{75} and Glu^{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^{2+}-binding loop (HNIEVLEGNEQ). The resulting "global" total enthalpies (ΔH) 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^{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 HNIEVLEGNEQ shows the positioning of substrate residues relative to the electrostatic features of the CTRC surface (Fig. 4*C*); residues Asn^{77}-Ile^{78}-Glu^{79}-Val^{80}-Leu^{81}-Glu^{82}-Gly^{83}-Asn^{84}-Glu^{85} fill the largely hydrophobic cleft between the flanking clusters of positive charge, with the Leu^{81} side chain embedded in the hydrophobic S_1 subsite. Surprisingly, none of the acidic side chains of the substrate, Glu^{79}, Glu^{82}, or Glu^{85}, which fill the P_3, P_1', and P_4' positions, respectively, form direct salt bridges with the clustered basic side chains of CTRC in the energy minimized docked model (Fig. 5*A*). Instead, it would appear that the complex stabilization attributable to charge complementarity derives from longer range electrostatic interactions. Trypsinogen Glu^{82} in the P_1' position is stabilized by CTRC Arg^{62A} via an interaction bridged by H-bonds with the P_3' side chain of trypsinogen Asn^{84}. The trypsinogen Glu^{85} P_4' side chain is positioned equidistant from the guanidinium groups of CTRC Arg^{39} and Arg^{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. 5*A*).

The docked model of the competing substrate cleavage site, APFDDDDK (the trypsinogen activation peptide), shows that this shorter substrate fills only the S_1-S_3 subsites on the non-primed side of the cleft, forming fewer H-bonds and hydrophobic interactions as compared with the longer substrate (Fig. 5*B*). In this model, Phe^{18} fills the hydrophobic S_1 subsite, and the

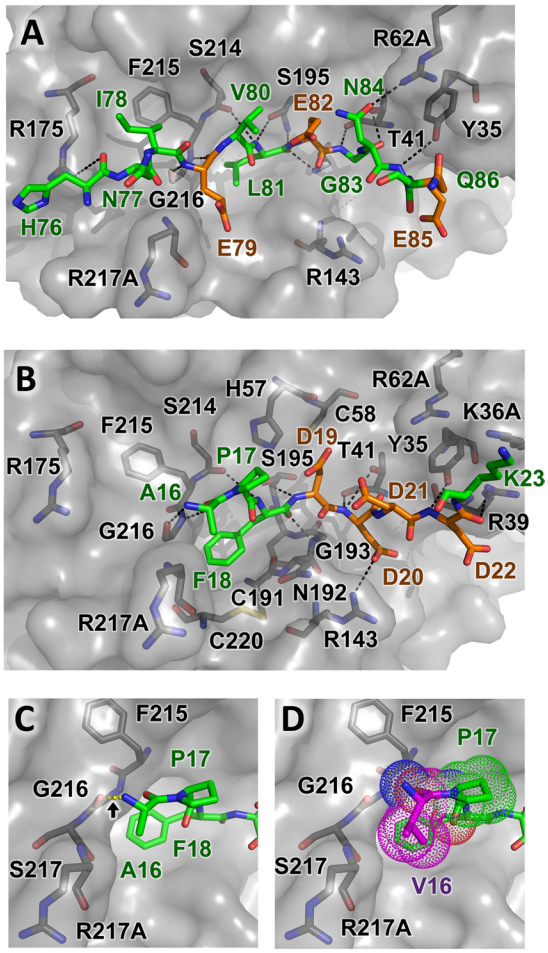

FIGURE 5. **Structural modeling of CTRC-substrate interactions.** *A*, trypsinogen Ca^{2+}-binding loop substrate positioning relative to nearby CTRC residues illustrates H-bonds formed between enzyme and substrate as *black dotted lines*. None of substrate acidic residues Glu^{79}, Glu^{82}, or Glu^{85} lie in close enough proximity to the CTRC basic side chains to form direct salt bridges. *B*, a similar view showing a model of trypsinogen activation peptide cleavage sequence bound to CTRC. This shorter alternative substrate forms one fewer non-primed side hydrogen bond with the enzyme, and lacks the hydrophobic stabilization that would be provided by a P_4 substrate residue. However, a direct salt bridge is formed between P_2' residue Asp^{20} and CTRC Arg^{143}. *C* and *D*, minimal differences are present in models of trypsinogen activation peptide (*C*) and p.A16V mutant (*D*) bound to CTRC. Where the Ala^{16} N-terminal amine is predicted to hydrogen bond with the CTRC Gly^{216} carbonyl (*C*), rotation of the Val^{16} (*magenta*) eliminates this bond but the bulkier side chain forms additional hydrophobic contacts with the Pro^{17} ring and the β-carbon of Arg^{217A} (*D*).

substrate does form a single direct salt bridge between Asp^{20} at the P_2' position and Arg^{143} (Arg^{162}) of CTRC (Fig. 5*B*). However, as was the case for the Ca^{2+}-binding loop site, electrostatic stabilization conferred by the Asp residues at the P_1', P_3', and P_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^{16} at the P_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-

DDDK very closely resembles that of the wild-type sequence, with significant differences confined to the mutated residue itself. The N-terminal amine of Ala^{16} H-bonds to the CTRC Gly^{216} carbonyl in the model of the wild-type sequence (Fig. 5*C*), whereas in the mutant model this H-bond is disrupted by slight rotation of Val^{16} to optimize hydrophobic contacts with the Pro^{17} ring and with the β-carbon of Arg^{217A} (Fig. 5*D*). 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 Ca^{2+}-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.

Nonspecific 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 P_1 "anchor" residue within the S_1 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 P_1 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 P_1 Leu in the trypsinogen Ca^{2+}-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 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 Ca^{2+}-binding loop.

Our modeling results are consistent with the idea that when sterically accessible, the 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 Ca^{2+}, the 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 β-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 both residues must conform to the idealized binding mode of a canonical loop, in which the P_3 residue main chain assumes an antiparallel β-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., Tóth, M., Mössner, J., and Sahin-Tóth, M. (2006) Mutations of human cationic trypsinogen (PRSS1) and chronic pancreatitis. *Hum. Mutat.* **27,** 721–730
4. Szabó, A., and Sahin-Tóth, 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-Tóth, M. (2006) Chymotrypsin C (caldecrin) stimulates autoactivation of human cationic trypsinogen. *J. Biol. Chem.* **281,** 11879–11886
6. Szmola, R., and Sahin-Tóth, M. (2007) Chymotrypsin C (caldecrin) 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., Rebours, V., Delhaye, M. N., Andrén-Sandberg, A., Weiss, F. U., Mountford, R., Harcus, M. J., Niemczyck, E., Vitone, L. J., Dodd, S., Jørgensen, M. T., Ammann, R. W., Schaffalitzky de Muckadell, O., Butler, J. V., Burgess, P., Kerr, B., Charnley, R., Sutton, R., Raraty, M. G., Devière, J., Whitcomb, D. C., Neoptolemos, J. P., Lévy, P., Lerch, M. M., Greenhalf, 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., Bhatia, E., Ózsvári, B., Landt, O., Schulz, H. U., Gress, T. M., Pfützer, R., Löhr, 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., Büning, C., Ockenga, J., Kage, A., Groneberg, D. A., Nickel, R., Berg, T., Wiedenmann, B., Bödeker, H., Keim, V., Mössner, J., Teich, N., and Sahin-Tóth, 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 Férec, C. (2008) Association of rare chymotrypsinogen C (CTRC) gene variations in patients with idiopathic chronic pancreatitis. *Hum. Genet.* **123,** 83–91
11. Zhou, J., and Sahin-Tóth, 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-Tóth, 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. J., and Craik, C. S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4,** 337–360

15. Szmola, R., Bence, M., Carpentieri, A., Szabó, A., Costello, C. E., Samuelson, J., and Sahin-Tóth, M. (2011) Chymotrypsin C is a co-activator of human pancreatic procarboxypeptidases A1 and A2. *J. Biol. Chem.* **286,** 1819–1827
16. Szabó, A., and Sahin-Tóth, M. (2012) Determinants of chymotrypsin C cleavage specificity in the calcium-binding loop of human cationic trypsinogen. *FEBS J.* **279,** 4283–4292
17. Szabó, A., Héja, D., Szakács, D., Zboray, K., Kékesi, K. A., Radisky, E. S., Sahin-Tóth, M., and Pál, G. (2011) High affinity small protein inhibitors of human chymotrypsin C (CTRC) selected by phage display reveal unusual preference for P4′ acidic residues. *J. Biol. Chem.* **286,** 22535–22545
18. Komiyama, T., and Fuller, R. S. (2000) Engineered eglin c variants inhibit yeast and human proprotein processing proteases, Kex2 and furin. *Biochemistry* **39,** 15156–15165
19. Otwinowski, Z., and Minor, W. (1997) in *Macromolecular Crystallography, Part A* (Carter, C. W., Jr., and Sweet, R. M., eds) pp. 307–326, Academic Press, New York
20. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61,** 458–464
21. Huang, K., Strynadka, N. C., Bernard, V. D., Peanasky, R. J., and James, M. N. (1994) The molecular structure of the complex of *Ascaris* chymotrypsin/elastase inhibitor with porcine elastase. *Structure* **2,** 679–689
22. Frigerio, F., Coda, A., Pugliese, L., Lionetti, C., Menegatti, E., Amiconi, G., Schnebli, H. P., Ascenzi, P., and Bolognesi, M. (1992) Crystal and molecular structure of the bovine α-chymotrypsin-eglin c complex at 2.0-Å resolution. *J. Mol. Biol.* **225,** 107–123
23. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) ARP/wARP and molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1445–1450
24. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6,** 458–463
25. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
26. Emsley, P., and Cowtan, K. (2004) Coot. Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
27. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace. A web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201
28. Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values. *J. Chem. Theory Comput.* **7,** 2284–2295
29. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pK29a values. *Proteins* **61,** 704–721
30. Polak, E., and Ribière, G. (1969) Note sur la convergence de méthodes de directions conjuguées. *Rev. Fr. d'informatique et de Recherche Opérationnelle* **16,** 35–43
31. Caulfield, T., and Medina-Franco, J. L. (2011) Molecular dynamics simulations of human DNA methyltransferase 3B with selective inhibitor nanaomycin A. *J. Struct. Biol.* **176,** 185–191
32. Caulfield, T., and Devkota, B. (2012) Motion of transfer RNA from the A/T state into the A-site using docking and simulations. *Proteins* **80,** 2489–2500
33. Loving, K., Salam, N. K., and Sherman, W. (2009) Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J. Comput. Aided Mol. Des.* **23,** 541–554
34. Salam, N. K., Nuti, R., and Sherman, W. (2009) Novel method for generating structure-based pharmacophores using energetic analysis. *J. Chem. Inf. Model* **49,** 2356–2368
35. Gomis-Rüth, F. X., Gómez, M., Bode, W., Huber, R., and Avilés, F. X. (1995) The three-dimensional structure of the native ternary complex of bovine pancreatic procarboxypeptidase A with proproteinase E and chymotrypsinogen C. *EMBO J.* **14,** 4387–4394
36. Birktoft, J. J., Kraut, J., and Freer, S. T. (1976) A detailed structural comparison between the charge relay system in chymotrypsinogen and in α-chymotrypsin. *Biochemistry* **15,** 4481–4485
37. Page, M. J., and Di Cera, E. (2008) Serine peptidases. Classification, structure and function. *Cell. Mol. Life Sci.* **65,** 1220–1236
38. Seemüller, U., Meier, M., Ohlsson, K., Müller, H. P., and Fritz, H. (1977) Isolation and characterization of a low molecular weight inhibitor (of chymotrypsin and human granulocytic elastase and cathepsin G) from leeches. *Hoppe-Seylers Z. Physiol. Chem.* **358,** 1105–1107
39. Bode, W., Papamokos, E., Musil, D., Seemueller, U., and Fritz, H. (1986) Refined 1.2 A crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin. *EMBO J.* **5,** 813–818
40. McPhalen, C. A., Schnebli, H. P., and James, M. N. (1985) Crystal and molecular structure of the inhibitor eglin from leeches in complex with subtilisin Carlsberg. *FEBS Lett.* **188,** 55–58
41. McPhalen, C. A., and James, M. N. (1988) Structural comparison of two serine proteinase-protein inhibitor complexes. Eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* **27,** 6582–6598
42. McPhalen, C. A., Svendsen, I., Jonassen, I., and James, M. N. (1985) Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin Novo. *Proc. Natl. Acad. Sci. U.S.A.* **82,** 7242–7246
43. Bode, W., and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204,** 433–451
44. Krowarsch, D., Cierpicki, T., Jelen, F., and Otlewski, J. (2003) Canonical protein inhibitors of serine proteases. *Cell Mol. Life Sci.* **60,** 2427–2444
45. Radisky, E. S., and Koshland, D. E., Jr. (2002) A clogged gutter mechanism for protease inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 10316–10321
46. Kardos, J., Bódi, A., Závodszky, P., Venekei, I., and Gráf, L. (1999) Disulfide-linked propeptides stabilize the structure of zymogen and mature pancreatic serine proteases. *Biochemistry* **38,** 12248–12257
47. Wang, H., Yuan, D., Xu, R., and Chi, C. W. (2011) Purification, cDNA cloning, and recombinant expression of chymotrypsin C from porcine pancreas. *Acta Biochim. Biophys. Sin.* **43,** 568–575
48. Bence, M., and Sahin-Tóth, M. (2011) Asparagine-linked glycosylation of human chymotrypsin C is required for folding and secretion but not for enzyme activity. *FEBS J.* **278,** 4338–4350
49. Otzen, D. E., Rheinnecker, M., and Fersht, A. R. (1995) Structural factors contributing to the hydrophobic effect. The partly exposed hydrophobic minicore in chymotrypsin inhibitor 2. *Biochemistry* **34,** 13051–13058
50. Heinz, D. W., Hyberts, S. G., Peng, J. W., Priestle, J. P., Wagner, G., and Grütter, M. G. (1992) Changing the inhibitory specificity and function of the proteinase inhibitor eglin c by site-directed mutagenesis. Functional and structural investigation. *Biochemistry* **31,** 8755–8766
51. Lu, W., Randal, M., Kossiakoff, A., and Kent, S. B. (1999) Probing intermolecular backbone H-bonding in serine proteinase-protein inhibitor complexes. *Chem. Biol.* **6,** 419–427
52. Lu, W. Y., Starovasnik, M. A., Dwyer, J. J., Kossiakoff, A. A., Kent, S. B., and Lu, W. (2000) Deciphering the role of the electrostatic interactions involving Gly^{70} in eglin C by total chemical protein synthesis. *Biochemistry* **39,** 3575–3584
53. Radisky, E. S., King, D. S., Kwan, G., and Koshland, D. E., Jr. (2003) The role of the protein core in the inhibitory power of the classic serine protease inhibitor, chymotrypsin inhibitor 2. *Biochemistry* **42,** 6484–6492
54. Radisky, E. S., Lu, C. J., Kwan, G., and Koshland, D. E., Jr. (2005) Role of the intramolecular hydrogen bond network in the inhibitory power of chymotrypsin inhibitor 2. *Biochemistry* **44,** 6823–6830
55. Salameh, M. A., Soares, A. S., Hockla, A., and Radisky, E. S. (2008) Structural basis for accelerated cleavage of bovine pancreatic trypsin inhibitor (BPTI) by human mesotrypsin. *J. Biol. Chem.* **283,** 4115–4123
56. Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.* **220,** 507–530
57. Rajamani, D., Thiel, S., Vajda, S., and Camacho, C. J. (2004) Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 11287–11292
58. Schreiber, G., Haran, G., and Zhou, H. X. (2009) Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* **109,** 839–860
59. Grünberg, R., Leckner, J., and Nilges, M. (2004) Complementarity of structure ensembles in protein-protein binding. *Structure* **12,**

2125–2136
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