Structure
Article

Structural Insight into the Bacterial Mucinase StcE Essential to Adhesion and Immune Evasion during Enterohemorrhagic *E. coli* Infection

Angel C.Y. Yu,[1] Liam J. Worrall,[1] and Natalie C.J. Strynadka[1,*]
[1]Department of Biochemistry and Molecular Biology and the Center for Blood Research, Life Sciences Center, 2350 Health Sciences Mall, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
*Correspondence: natalie@byron.biochem.ubc.ca
DOI 10.1016/j.str.2012.02.015

SUMMARY

Mucin glycoproteins with large numbers of O-linked glycosylations comprise the mucosal barrier lining the mammalian gastrointestinal tract from mouth to gut. A critical biological function of mucins is to protect the underlying epithelium from infection. Enterohemorrhagic *Escherichia coli* (EHEC), the mediator of severe food- and water-borne disease, can breach this barrier and adhere to intestinal cells. StcE, a ~100 kDa metalloprotease secreted by EHEC, plays a pivotal role in remodeling the mucosal lining during infection. To obtain mechanistic insight into its function, we have determined the structure of StcE. Our data reveal a dynamic, multidomain architecture featuring an unusually large substrate-binding cleft and a prominent polarized surface charge distribution highly suggestive of an electrostatic role in substrate targeting. The observation of key conserved motifs in the active site allows us to propose the structural basis for the specific recognition of α-O-glycan-containing substrates. Complementary biochemical analysis provides further insight into its distinct substrate specificity and binding stoichiometry.

INRODUCTION

Mucin glycoproteins in the mucosal lining of the human digestive tract pose a significant challenge for potential pathogens that come into contact with the surface. They spatially separate the host epithelial cells from microorganisms in the lumen and consist of an adherent inner layer and a dynamic outer layer that is constantly sloughed off and renewed (reviewed in Linden et al., 2008). Ingested microbes must first avoid agglutination by salivary mucins and removal by mucociliary action, as well as penetrate through a thick mucus layer masking the epithelial cells before colonization can take place. With thickness ranging from 70–100 μm in salivary films to 800–900 μm in the colon, mucins form a protective physical barrier for the underlying epithelium (reviewed in Derrien et al., 2010). Mucins are densely glycosylated proteins characterized by Pro/Thr/Ser (PTS)-rich domains. The hydroxyls of Ser or Thr residues therein can be modified by various O-linked glycans through the linking sugar α-N-acetylgalactosamine (GalNAc) (reviewed in Hang and Bertozzi, 2005) and the carbohydrate chains often occur in clusters along the PTS tandem repeats. The high density of oligosaccharides projecting from the peptide backbone results in an extended, "bottle brush"-like structure with dimensions 2.5 to 3 times greater than unglycosylated, denatured globular proteins with similar number of residues (reviewed in Jentoft, 1990). Electrostatic repulsion and steric exclusion from the negatively charged O-linked sugars confer protease resistance to the mucin peptide backbone (reviewed in Van den Steen et al., 1998) and, in turn, impede access to the host epithelium underneath.

Enterohemorrhagic *Escherichia coli* of serotype O157:H7 (EHEC) is a human pathogen associated with food-borne infections and gastrointestinal disease. These highly infectious bacteria colonize epithelial cells in the colon and can cause hemolytic uremic syndrome with a potentially lethal outcome (reviewed in Nataro and Kaper, 1998). EHEC has the ability to penetrate the mucosal lining and adhere intimately to eukaryotic host cells but the molecular mechanisms of this process, which is required for both colonization and infection, are only partially understood. StcE (sercreted protease of C1-esterase inhibitor), which EHEC secretes via the type II general secretory pathway during infection, has been implicated to assist the penetration step (Lathem et al., 2002; Paton and Paton, 2002). Its expression from the pO157 virulence plasmid is activated by Ler, a global regulator of various genes essential for EHEC pathogenesis, including the LEE (locus of enterocyte effacement)-encoded type III secretion system and its translocated effectors. Indeed, Grys et al. (2005) demonstrated that StcE played a critical role in colonization by proteolytically unmasking the host cell surface to allow intimate adherence of EHEC to gut epithelial cells. The decreased virulence of the *Aeromonas hydrophila tagA* mutant (64% protein sequence identity with StcE) in a mouse model of infection also supports its proposed function as a host surface-clearing mucinase (Pillai et al., 2006). It was further concluded that colonization defects of an EHEC type II secretion knockout might be related to StcE deficiency (Ho et al., 2008).

The role of StcE in virulence is attributed to its zinc metalloprotease mediated mucinase activity toward specific mucin-type glycoproteins. It degrades and reduces the viscosity of the mucus layer via cleavage of mucin 7 and glycoprotein 340 present in saliva and other tissues (Grys et al., 2005). These targets

Table 1. Data Collection and Refinement Statistics

	StcE SeMet Derivative	StcE H132-N251
Data Collection		
Space group	$C222_1$	H3
Mol/au	1	1
Cell dimensions		
a, b, c (Å)	76.5, 186.0, 188.8	56.0, 56.0, 96.9
α, β, γ (Å)	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution (Å)	2.5 (2.56–2.50)[a]	1.6 (1.69–1.60)[a]
Total reflections	275,104	80,354
Unique reflections	47,052	14,661
R_{pim}[b]	5.5 (36.9)	2.1 (11.3)
I/σI	9.2 (1.8)	30.3 (7.3)
Completeness (%)	100 (100)	98.9 (92.8)
Redundancy	5.8 (5.9)	5.5 (5.1)
Refinement		
R_{work}/Rf_{ree}	0.21/0.25	0.18/0.21
B factor (Å^2)		
Protein	83	18.0
Ion	53	21.7 (7 iodides)
	63 (175 molecules)	26.0 (103 molecules)
Rmsds		
Bond length (Å)	0.010	0.020
Bond angles (°)	1.200	1.486
Ramachandran plot		
Favored (%)	95	98
Disallowed (%)	0.4	0

[a]Multiplicity-weighted R_{merge}.
[b]Values in parentheses are for the highest-resolution shell.

form part of the innate immune system by acting as receptor decoys for microbial adhesins to facilitate the clearance of bacteria-protein aggregates from the host. Thus, destruction of these defense molecules by StcE would allow EHEC to overcome the mucosal barrier for necessary contact with the epithelium to establish infection.

In addition to its role in mediating host cell colonization, StcE is believed to contribute to immune evasion. Once attached to colon epithelial cells, EHEC secretes various effector proteins that lead to lesions in the endothelium and intestinal lining, and the infiltration of leukocytes and complement proteins to the gut lumen (reviewed in Nataro and Kaper, 1998). StcE specifically targets leukocyte surface glycoproteins CD45 and CD43, which also contain glycosylation of the mucin type, and this provides a further competitive advantage to the pathogen and allows it to modulate the host immune system (Szabady et al., 2009). In particular, the loss of these sialoglycoproteins due to StcE proteolytic activity prevents immune cells from moving to the sites of infection and mounting the appropriate inflammatory response. Indeed, StcE-treated neutrophils exhibited poor chemotactic mobility. Notably, StcE also targets the complement system, a set of proteolytically regulated plasma proteins that form the central component of host innate immunity and mark pathogens for destruction (reviewed in Lambris et al., 2008). StcE inhibits proteolytic activation of the complement cascades by recruiting human C1-esterase inhibitor (C1-INH) to the cell surface (Lathem et al., 2004). C1-INH is a member of the serpin (serine protease inhibitor) family and an essential regulator of serine proteases in the complement pathways. Through specific interactions with the mucin-like N terminus of C1-INH (Figure S1A available online), StcE localizes the complement regulator to the cell membrane, where the C-terminal serpin domain is then in proximity to inactivate downstream components of the complement cascades. As a result of the inhibition, it was shown that serum-sensitive *E. coli*-K12 became more resistant to complement-mediated cell lysis in the presence of StcE-treated C1-INH (Lathem et al., 2004).

StcE belongs to the family M66 of zinc metalloproteases (Rawlings et al., 2010), which includes ToxR-activated gene A (TagA) from the *Vibrio cholera* pathogenicity island, and for which no structural information of any member is currently known. To obtain mechanistic insight into StcE function and to understand more fully how it subverts multiple mucosal defenses and the complement system of innate immunity, we have carried out X-ray crystallographic analysis of StcE and characterized its interaction with its substrate C1-INH using complementary biochemical and biophysical approaches.

RESULTS

Overall Structure

During EHEC infection, StcE is transported to the extracellular milieu via the type II general secretory pathway. For our structural studies, we were able to express sufficient quantities of the mature form of StcE devoid of its signal sequence (StcEΔ35, referred to hereafter as full-length StcE containing residues 36–898). Static light scattering analysis showed that full-length StcE is monomeric and monodisperse with a measured molecular mass (~100 kDa) close to the expected value (Figure S1B). After numerous attempts, we managed to obtain diffraction-quality crystals of a catalytically inactive (E447D) and surface-engineered mutant (K318A/K320A/E321A) of StcE and solved the structure by the SAD method, using crystals of selenomethionine-incorporated protein (Table 1). Examination of the engineered protein surface shows residues 318, 320, and 321 are distant from the active site. The alanine mutations at these positions allowed close packing by Tyr859 from a symmetry-related StcE molecule and this likely facilitated crystal contact formation.

The structure of apo-StcE, refined to a resolution of 2.5 Å, reveals a dynamic, modular architecture with three distinct globular domains, which we designate as IG, M, and C (Figures 1A–1C). These three domains are packed against each other and result in an overall T-shape, with IG and M forming the arms and C the body. The N-terminal domain IG adopts an immunoglobulin-like fold that consists of complementary segments spanning residues 57–122 and 257–287. Within domain IG, a sequence-variable region with weak electron density, which we designate as INS (insertion residues Lys150-Ser245), inserts between strands 5 and 6 of the immunoglobulin β sandwich (Figures 1A and 1B; Figure S2 sequence alignment). Immediately following domain IG is the metalloprotease catalytic module, domain M (Glu296-Val550). M is cradled in part by domain IG and visible

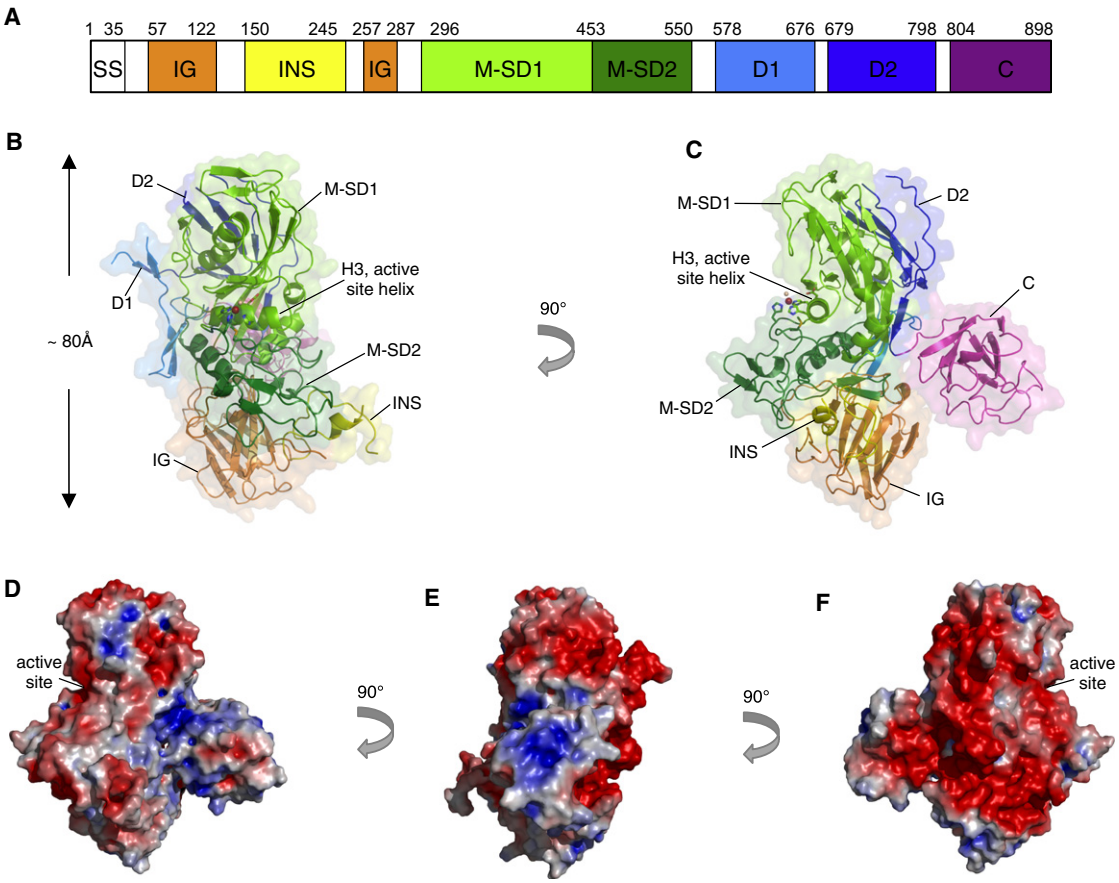

Figure 1. StcE Overall Structure

(A–D) (A) Domain boundaries observed in the StcE structure. The two subdomains of the metalloprotease module (M-SD1 and M-SD2) are colored in green, immunoglobulin-like domain (IG) in orange, insertion region (INS) in yellow, C-terminal domain (C) in purple and remaining disordered regions (D) in blue. SS stands for signal sequence. (B) Secondary structure and surface representation of StcE. The color scheme matches that in (A). Catalytic residues are shown in stick form, zinc and nucleophilic water as red and beige spheres, respectively. (C) Ninety degree rotation of the molecule in (B). (D) Electrostatic surface representation of StcE in the same orientation as in (C), colored from −4 kT/e (red, negative) to +4 kT/e (blue, positive) using APBS (Baker et al., 2001).
(E and F) The molecule in (D) is rotated 90° and 180°, respectively.

portions of INS, resulting in a large buried surface area of ~3,955 Å^2 (Figures 1B and 1C). A partially resolved region, residues 679–798 (D2), docks against M on the opposite face of the active site. Viewing toward the active site, the D2 region and domains M and IG are approximately coplanar with a long axis of ~80 Å (Figure 1B). Extending in opposite directions away from this axis are disordered structural elements, INS and D1, spanning Arg578-Arg676 (as judged by residual density and weak anomalous signals likely reflecting SeMet594, 598, and 608 in region D1). This unique arrangement of distinct, noncatalytic domains observed around the metalloprotease domain likely allows for additional regulatory functions including substrate recognition. The chain terminates in domain C (Glu804-Lys898), which is tethered to D2 through a short linker and adopts a β/γ crystalline fold. This domain is located ~180° from the active site and makes minimal contact with rest of the protein.

The electrostatic surface potential of StcE calculated based on our crystal structure reveals a highly polarized charge distribution for the metalloprotease (Figures 1D–1F). Specifically, a large electronegative surface exists on one face of the molecule, contiguous along the three domains IG, M, and C. In contrast, the opposite plane is primarily electropositive and hydrophobic. We suggest the uneven charge distribution likely plays a role in substrate targeting at the host-pathogen interface. This is reminiscent of the electrostatic features utilized by the carbohydrate-binding sialidase/trans-sialidase family. Charge repulsion from the electronegative face is believed to orient the more electropositive substrate-binding site toward the negatively charged glycoconjugate substrates and to facilitate reversible binding (Luo et al., 1998).

Metalloprotease Domain

StcE possesses the extended zinc-binding motif HExxHxxGxxH within its catalytic domain (Glu296-Val550), the boundaries of which were ambiguous from sequence analysis and are now clearly defined in our crystal structure. Overall, domain M adopts a characteristic mixed α+β fold and a conserved Met-turn (StcE Met518), features that place it within the metzincin superfamily of metalloproteases including astacins, ADAMs (a disintegrin and

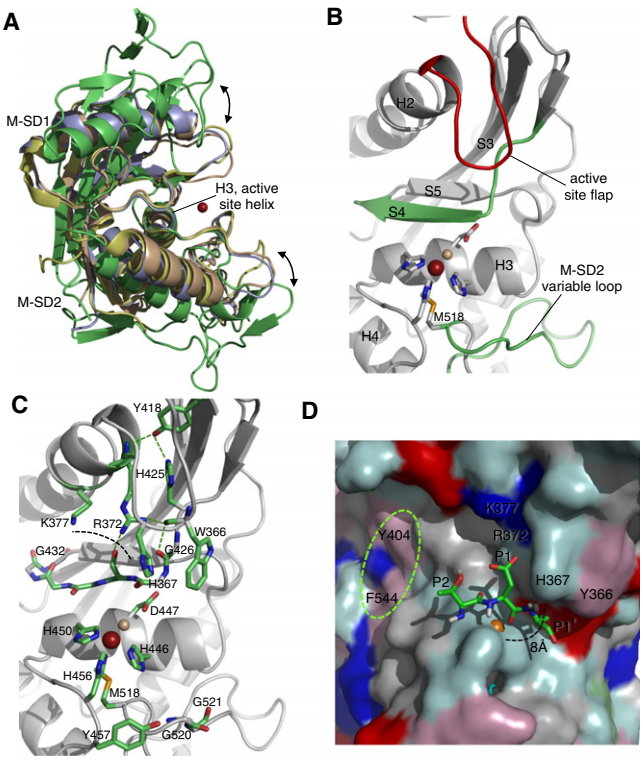

Figure 2. StcE Active Site

(A) Superposition of the StcE metalloprotease domain, in green, with related metzincins (yellow, PDB 1IAG with a Z-score of 6.4; wheat, 2ERO, 6.5; blue, 2DW2, 6.8) along the active site helix. The wider angular opening between M-SD1 and M-SD2 in StcE is believed to accommodate large, densely O-glycosylated substrates. The catalytic zinc ion is shown as a red sphere.

(B) Structural elements known to mediate substrate binding in metzincins are highlighted in green. The active site flap in red is unique to StcE. Note the high curvature in the strand preceding S4. Catalytic residues are shown in stick form, zinc as a red sphere, and nucleophilic water as a beige sphere.

(C) Orientation of StcE is identical to the molecule in (B). Conserved structural motifs potentially involved in substrate binding and transition state stabilization surround the catalytic residues. Black dashed line delineates the proposed S1 pocket. Green dashed lines represent hydrogen bonds.

(D) Docking of the C1-INH tripeptide, $\mathbf{T^{111}DS}$, in the active site. StcE surface is colored by residue type, with cyan being polar; blue, basic; red, acidic; pink, aromatic and grey, hydrophobic. The catalytic zinc ion is shown as an orange sphere. Dashed green line outlines the proposed sugar-binding site.

See also Figure S3.

metalloprotease), and matrixins (reviewed in Gomis-Rüth, 2009). Notably, a central substrate-binding cleft containing the active site separates this domain into two subdomains M-SD1 and M-SD2 (Figures 1C and 2A). M-SD1 consists of a twisted β sheet (order β6, β2, β1, β3, β5, β4) packed on both sides by three helices (H1–H3). A sharp turn introduced by the conserved Gly453 in the zinc-binding motif along the active site helix (H3) then leads to M-SD2, which, after a series of convoluted loops, terminates in H4. In addition to the metzincin-like topology, StcE contains unique structural elements that may have evolved to allow cleavage of densely O-glycosylated substrates. In related single-domain zinc metalloproteases, namely thermolysin and astacin, ligand-binding triggers closure of the ligand-free, open conformation via subdomain motion and reorients the active site residues for catalysis (Grams et al., 1996; Holland et al., 1992). Overlay of StcE with structurally homologous metzincins shows the former to have an active site cleft with a wider angular spacing in the analogous "open" conformation (Figure 2A). The relative orientation of M-SD1 and M-SD2 is ~10° wider in StcE (see Experimental Procedures), as is the spacing between corresponding elements known to recognize primed-side residues of the substrate via β strand-like hydrogen bonds. These include structural elements connecting S3 and S4, which exhibit unusually high curvature in StcE and a variable loop joining the conserved methionine (StcE Met518) and helix H4 in M-SD2 (Figure 2B). Furthermore, the M-SD2 variable region (Leu454-Thr534) is considerably more elaborate (in organization and length) than those typically observed in metzincins, forming a large, concave surface toward the catalytic center (Figures 1C and 2A). Collectively, these features may have evolved to facilitate approach by the expanded dimensions and "bottle brush"-like structures of densely glycosylated mucin-like substrates, including C1-INH, CD43, and CD45 (Cyster et al., 1991; reviewed in Holmes, 2006; Perkins et al., 1990).

Catalytic Site and Implications for Substrate Specificity

Although sequence identity is limited among metzincins, structural superposition of StcE with other members of the family reveals a number of conserved catalytic residues. The essential zinc ion is coordinated in the characteristic tetrahedral fashion by His446 and His450 on the active site helix (H3), His456, and a water molecule (Figure 2C). Met518 in the conserved Met-turn forms the hydrophobic base of the active site. In our crystal structure, Asp447 has been substituted for the Glu general base in wild-type (WT) StcE to facilitate crystallization. Being one methylene unit shorter, Asp447 is ~5 Å away from the observed nucleophilic water, thus consistent with the inability of this mutant to hydrolyze substrates (Lathem et al., 2002).

In all metzincin clan members structurally characterized to date, S4 in the twisted β sheet forms the upper rim of the active site cleft and binds primarily to the nonprimed end of the substrate in an antiparallel fashion (reviewed in Gomis-Rüth, 2009). In StcE, a stretch of conserved glycines, from Gly426 to Gly432 with an internal Ser428, occurs along S4 (Figures 2B and 2C). The poly-Gly motif likely confers plasticity to the active site cleft, allowing StcE to cleave substrates with somewhat lax specificity preferences. This structural feature may explain why it has been difficult to identify a consensus sequence among the glycosylated PTS-rich regions of StcE substrates (Szabady et al., 2011). Due to intramolecular interactions between the α-O-linked GalNAc and the peptide core (Coltart et al., 2002), the O-glycan clusters constrain the mucin-like peptide chain into a specific, extended conformation (reviewed in Jentoft, 1990). High conformational freedom of the poly-Gly backbone likely facilitates recognition of the O-glycan induced conformation. Indeed, the phi and psi angles of the well-ordered Gly426 and 430 lie in less favorable regions of the Ramachandran plot (data not shown). Immediately upstream of H2 is a flap containing the strictly conserved Trp366 and His367 (Figures 2B and 2C). Their side chains project into and constrict the active site, providing some level of subsite specificity. Importantly, the backbone orientation of this flap is held in position via hydrogen bonding by several invariant residues, namely, Gly426 of the

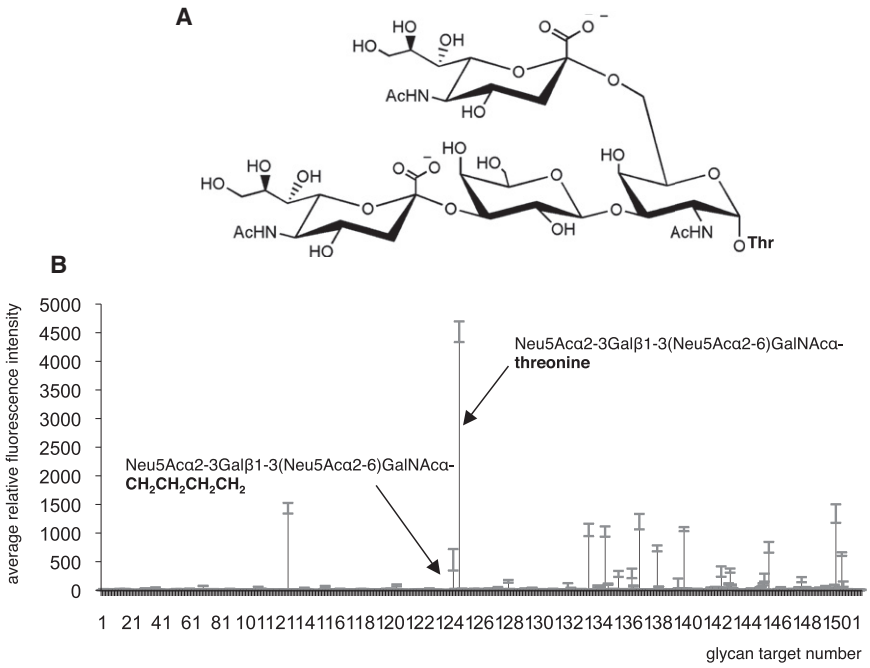

Figure 3. Mammalian Glycan Array Screening of StcEΔ35E447D

(A) Chemical structure of the highest affinity ligand, Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα-threonine.

(B) The strength of binding for each glycan array target represented on the x-axis is proportional to the measured fluorescent intensity (shown on the y-axis). Note the preferential binding of StcE to the O-glycan in (A) over a similar ligand substituted with the aliphatic linker -$CH_2CH_2CH_2CH_2$-. The error bar associated with each fluorescence signal measurement refers to the standard error of the mean.

poly-Gly strand, Arg372, and His425 side chains, which are further stabilized by the phenyl ring of Tyr418 (Figure 2C). By virtue of its proximity to the active site, His 367 may be involved in stabilizing the carboxyanionic transition state in addition to substrate binding. Tyr457 located downstream of the zinc ligand His456 could potentially fulfill a similar role (Figure 2C). The residue is strongly though not strictly conserved among M66 members (Figure S2). It may undergo a ''tyrosine switch'' or shift in position associated with substrate and/or transition state stabilization analogous to other family-specific nonzinc-binding tyrosines in this position (reviewed in Gomis-Rüth, 2003).

StcE/C1-Esterase Inhibitor Interaction

StcE has been shown to specifically require interaction with O-glycosylated regions of its substrates (Lathem et al., 2004; Szabady et al., 2009); however, little is known about the mode of interaction and substrate recognition sequence. We have chosen the commercially available C1-INH as our model substrate as it is more amenable to biochemical characterization than the membrane-associated mucin-type substrates CD43 and CD45. We were able to trap a complex of StcE bound to human C1-INH using the inactive E447D mutant. Static light analysis of the resulting metalloprotease/serpin complex revealed a 1:1 binding stoichiometry (Figures S1B–S1D). We tried to define the binding interface in more detail by X-ray crystallographic approaches. However, various attempts to obtain a high-resolution structure of the complex were unsuccessful, likely due to the inherent structural heterogeneity of O-glycosylated C1-INH.

To further probe the substrate specificity of StcE, we next sought to identify the cleavage site by N-terminal sequencing of the StcE-treated C1-INH fragment, which has an apparent molecular weight (60–65 kDa) comparable to that reported in the literature (Lathem et al., 2002). The resulting primed-side residues match closely to S^{113}PTQP in the published sequence except the T115L substitution (Figure S1A), possibly representing a sequence variant. The cleavage occurred within the O-glycosylated N-terminal region of C1-INH, in agreement with the fact that its C-terminal serpin domain remains active upon StcE binding (Lathem et al., 2004). N-terminal to the scissile bond, the P1 residue Asp112 of C1-INH could conceivably fit in a shallow pocket created by the invariant His367, Arg372, and Lys377 (Figure 2C). We attempted to fit the C1-INH tripeptide spanning P2-P1' (T^{111}DS) into the StcE active site by docking (Trott and Olson, 2010), and the accepted, energy-minimized conformation of the docked ligand shows a reasonable binding mode with respect to the surrounding catalytic residues (Figure 2D). To further identify potential carbohydrate moieties that StcE recognizes, we subjected the full-length protease to mammalian glycan array analysis. The screen revealed Neu5Acα2-3Galβ1-3(Neu5Acα2-6) GalNAcα-threonine as the top ligand (Figures 3A and 3B) (see Discussion).

Curiously, the poorly resolved density corresponding to the INS region (Lys150-Ser245) appears to approach the active site (Figure 1B). We generated a truncation mutant of this sequence-variable region (ΔINS) (Figure S2) to test its involvement in substrate binding. ΔINS exhibited decreased proteolytic activity toward C1-INH compared to WT StcE (Figure 4A) (circular dichroism spectrum confirmed no large perturbation to the overall structure due to the mutation, as shown in Figures 4B and 4C). This suggests the insertion may contribute to substrate binding and specificity although we cannot exclude the possibility of additional sequence determinants in the absence of a complete structure. Region INS has a propensity to form primarily β strands according to secondary structure prediction (Cole et al., 2008). The fact that we did not observe well-ordered density for INS in our model suggests this region may adopt multiple conformations that contribute to the ability of StcE to capture a variety of complex substrates at the cell surface. To explore this hypothesis further, we solved the structure of the isolated INS domain using an iodide soak SIRAS technique. Indeed, domain INS folds independently as a mixed, eight-stranded β sandwich with homology to the T4 bacteriophage structural protein gp9 (Z-score of 6.2) and contains an additional β strand followed by a perpendicular arrangement of

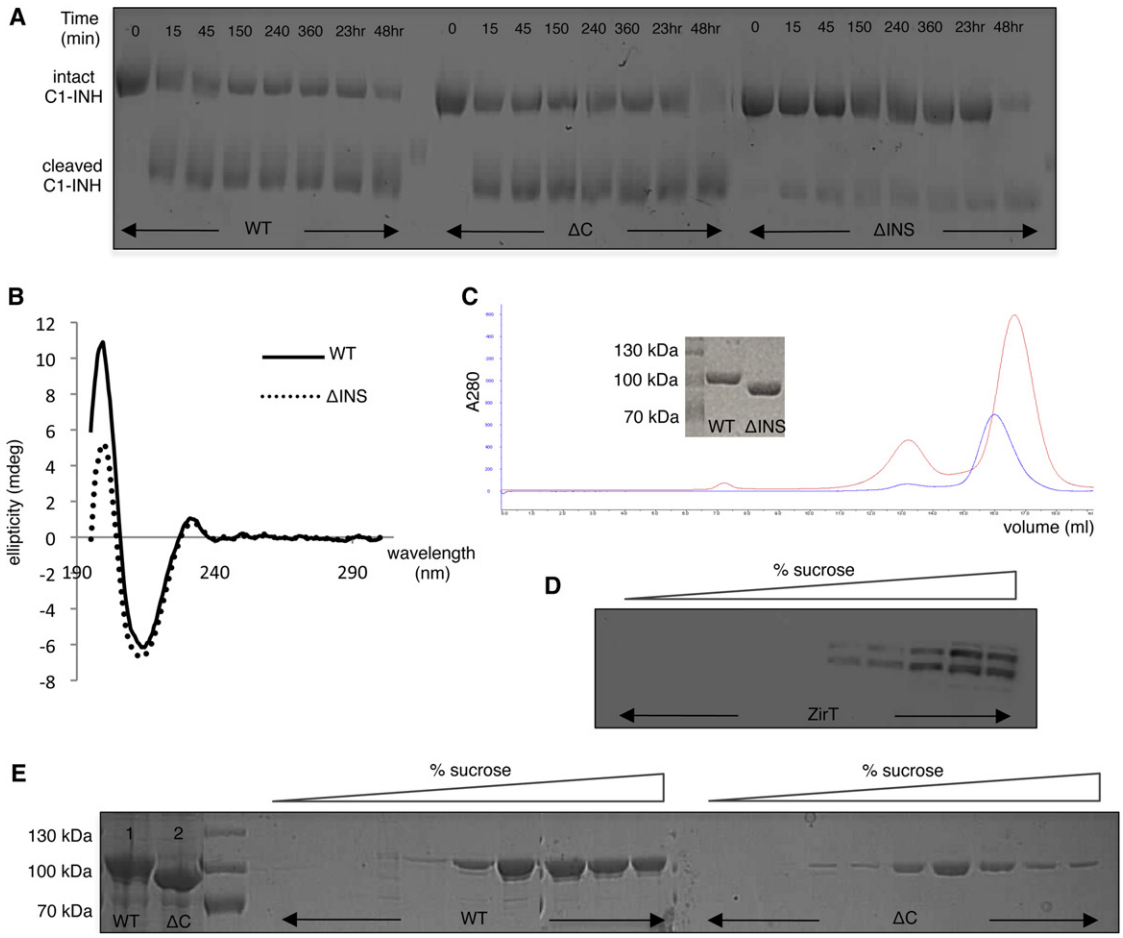

Figure 4. Biochemical Characterization of StcE Domain Variants

(A) Time-dependent proteolytic cleavage of C1-INH by WT StcE, C-terminal truncation (ΔC) and insertion deletion (ΔINS) mutants. Time was measured in minutes unless otherwise indicated (hour, hr). Relevant lanes were assembled from multiple SDS-PAGE gels.

(B) Circular dichroism analysis of WT StcE and ΔINS. The deletion mutant retained secondary structure elements; its impaired proteolytic efficiency was not due to misfolding.

(C) Gel filtration elution profiles of monomeric WT StcE (blue) and ΔINS (red). SDS-PAGE analysis of respective peak fractions is shown in the inset.

(D) Cellular localization analysis of the outer membrane protein control ZirT using increasing concentrations of sucrose (30% to 55%), as denoted by the gradient above the SDS-PAGE gel.

(E) Sucrose density gradient analysis of WT StcE and its C-terminal truncation mutant (StcEΔC) revealed a reduced amount of ΔC in the OMP fractions. The decrease was unlikely due to poor expression of StcEΔC as similar levels of WT StcE and the mutant were present in their respective total lysates in lanes 1 and 2. See also Figure S1.

two α helices in its N-terminal region. We then positioned the compact module in the context of the existing model to obtain a more complete description of the StcE structure and function (correct placement was confirmed by the anomalous signal of SeMet205). A long loop (His132-Lys150) links the N terminus of INS to domain IG and provides flexibility, placing the putative exosite domain ~35 Å away from the catalytic center in our model (Figure 5A). A set of electropositive surface patches and a conserved cluster of solvent-exposed Trp189, Arg191 and Tyr214, which appear suitable for glycoprotein binding, tilt toward the active site (Figures 5B and 5C). Being flexibly linked, domain INS may exist in various conformations to present a potential protein-binding interface relative to its respective targets.

Cell Surface Binding Features

StcE protects serum-sensitive *E. coli* from complement-mediated lysis by recruiting C1-INH to an opsonized, or antibody-coated cell surface, effectively increasing C1-INH local concentrations. The process requires direct interaction between the cells and StcE, as confirmed by flow cytometry analysis (Lathem et al., 2004), although the mechanism of surface localization is unclear. Structural homology search by DALI (Holm and Rosenström, 2010) suggests StcE's C-terminal domain may, in part, mediate such an interaction. The βγ-crystallin fold domain shows the highest similarity with protein S (Protein Data Bank [PDB] 1NPS, Z-score of 9.2 and rmsd of 2.4 Å over 77 aligned Cα), a surface coat protein that assembles around myxospores of *Myxococcus xanthus* during periods of cellular stress (Inouye

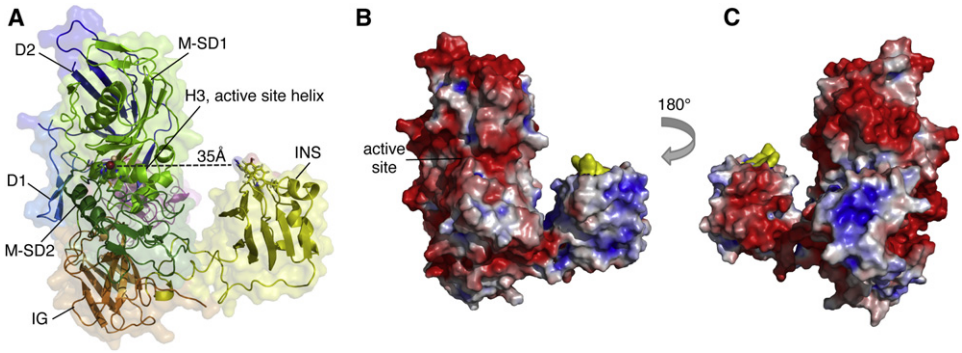

Figure 5. Relative Orientation of the Proposed Exosite Domain in StcE

(A) Position of domain INS in the overall StcE structure, shown in the same orientation as in Figure 1B. The conserved Trp189, Arg191, and Tyr214 potentially involved in carbohydrate binding are in yellow stick form.

(B) Electrostatic surface potential of the molecule in (A), colored from −4 kT/e (red, negative) to +4 kT/e (blue, positive). The conserved cluster containing solvent-accessible Trp189, Arg191, and Tyr214 is shaded in yellow.

(C) 180° rotation of the molecule in (B) reveals surface charge asymmetry in domain INS.

See also Figure S2.

et al., 1979). We evaluated the ability of the C-terminal domain to associate peripherally with the cell surface by comparing the cellular localization of WT StcE and its truncation mutant (ΔC) in an equilibrium sucrose density experiment. WT StcE cofractionated with the outer membrane protein (OMP) control ZirT (Gal-Mor et al., 2008) while significantly less of ΔC remained in the corresponding fractions (Figures 4D and 4E). Indeed, the calculated electrostatic surface of StcE reveals electropositive clusters and hydrophobic ridges in domain C that may provide complimentary charges for surface association (Figure 1E). Proteolytic efficiency of ΔC is comparable to the WT (Figure 4A), indicating the domain is dispensable for substrate binding.

DISCUSSION

StcE-like enzymes comprise the M66 metalloprotease family, for which only EHEC StcE and TagA orthologs from *A. hydrophila* (Pillai et al., 2006) and *V. cholera* have been characterized biochemically (Szabady et al., 2011), and no members structurally. As such, the StcE structure described here provides a foundation for understanding this family, including the molecular features that govern the critical mucinase action used to overcome the host innate immune response to bacterial infection.

We suggest the unusually open dimensions of the StcE substrate-binding cleft compared to other metzincins may reflect its unique ability to efficiently process large O-glycosylated substrates. In addition, the observed surface charge asymmetry in StcE likely has an impact on how the protease approaches its substrates. α-O-glycan clusters found along PTS-tandem repeats of mucin-type host glycoproteins impart an overall negative charge to the molecules. Of note, the predominately acidic face of StcE may provide charge repulsion that suitably orients the more electropositive target binding site toward surface localized substrates on the host cell, a strategy also adopted by members of the carbohydrate-binding sialidase/trans-sialidase family. We suggest these properties might have evolved to allow StcE to overcome the electrostatic shielding and general protease resistance afforded by O-glycosylated proteins (Garner et al., 2001). Measuring ~8 Å in width at the most narrow constriction, the StcE active site cleft widens further away from the catalytic center on both the primed and nonprimed ends (Figure 2D). This allows the protease to form optimal interactions at the scissile peptide bond while minimizing steric clashes with potential carbohydrate moieties linked to peptide residues nearby. These overall features, together with conserved structural motifs identified in the catalytic domain allow us to describe a probable mechanism for StcE's specificity toward mucin-type substrates.

When StcE is in the open conformation, O-glycoprotein substrates such as C1 esterase inhibitor (C1-INH) can access the clamp-like opening between M-SD1 and M-SD2 of the metalloprotease domain and be captured by StcE. A basic S1 specificity pocket formed by the conserved His367, Arg372, and Lys377 indicates a preference for acidic residues in this subsite of the enzyme active site. As shown by the interactions simulated by our characterized C1-INH peptide spanning P2-P1' (T^{111}DS), fitting of Asp112 at P1 is sterically and electrostatically compatible (Figure 2D). The bulky side chains of Trp366 and His367 in the conserved G^{364}GWHSG flap partially shield the active site from bulk solvent, increasing the local effective charge, particularly for electrostatic interactions in the S1 subsite (Figures 2C and 2D). Sequence analysis of C1-INH identified 14 potential O-glycosylation sites (Bock et al., 1986), including Thr111, which would fit as P2. Although the sugar moiety attached to Thr111 has yet to be verified experimentally, aromatic side chains of the nearby Tyr404 and Phe544 can potentially mediate carbohydrate binding through stacking interactions, a common theme in sugar binding (Figure 2D).

Our mammalian glycan microarray analysis of StcE identified the disialylated core I O-glycan Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα-threonine as the highest affinity ligand (Figure 3A). Previous lectin-binding assays have also confirmed the presence of similarly linked sugars on C1-INH (Schoenberger, 1992). Interestingly, substitution of the threonine in this glycoconjugate with the extended -$CH_2CH_2CH_2CH_2$- linker abrogates binding (Figure 3B), suggesting ligand recognition involves both the peptide and sugar moiety. Welch et al.

conducted a similar glycan array study and obtained different results that were reported to be inconclusive (available in a searchable public database on the Center for Functional Glycomics website http://www.functionalglycomics.org/static/consortium/links2Website.shtml). We acknowledge that we used the full-length StcE for our screen instead of a truncated version used in the other study and hypothesize that ligand binding requires cooperative interaction between the different StcE modules observed in our structure. Moreover, conservation of the poly-Gly motif in StcE, as well as Gly520 and Gly521 on the rims of the active site cleft indicates that conformational flexibility is essential for catalysis (Figures 2B and 2C). The glycine backbone can assume a wider range of torsional angles without undue strain, allowing the active site to adapt to and recognize specific α-O-glycan induced conformations in mucin-type glycoprotein substrates. Indeed, evidence from high-resolution NMR analysis of mucin glycopeptides based on the StcE substrate CD43, revealed a stable conformation arising from the interaction of α-O-GalNAc with adjoining peptide functional groups (Coltart et al., 2002). Specifically, the GalNAc amide proton hydrogen bonds to the glycosidic oxygen and the backbone carbonyl of the attached amino acid, while the hydrophobic face of the N-acetyl methyl forms medium-range interactions with methyl groups of the peptide side chain two positions downstream from the site of O-glycosylation. (Hashimoto et al., 2011).

While the electrostatic profile and open conformation of the StcE active site facilitates substrate approach, we speculate that substrate binding could trigger domain closure through induced-fit, leading to favorable protease-substrate interactions. Mutagenesis studies on the related thermolysin-like proteases (TLP) suggest that conserved glycines are involved in hinge-bending motions associated with closure and opening of the binding cleft during catalysis (Veltman et al., 1998). Interestingly, Gly441 of StcE lies in a structurally analogous position at the N-terminal end of the interdomain active site helix as the conserved Gly136 of TLP (Figure S2). The residue may mediate a similar hinge-like movement that optimally orients StcE and the target peptide toward one another.

Orientation of the docked C1-INH peptide in the StcE active site, with the P1 carbonyl within binding distance of the metal, allows for nucleophilic attack by the zinc-bound water on the *re*-face of the scissile peptide bond (Figure 2D). This is consistent with the catalytic register and substrate-binding mode of similar metzincins (reviewed in Gomis-Rüth, 2009). His367 and/or the nonligandingTyr457 could potentially stabilize the transient negative charge on the tetrahedral intermediate, contributing to the charge neutralization effect of the catalytic zinc cation nearby. This likely requires reorientation of the Tyr457-containing loop and rotation of the residue side chain to bring the phenolic hydroxyl into proximity, in a "tyrosine switch" motion observed in the related snapalysins (Kurisu et al., 1997). After hydrolysis, the cleaved target is liberated from the active site and the protease regenerated.

Despite extensive conserved surface features in the active site (Figure S3) and their shared specificity toward mucin-type glycoproteins, substrate profiles of StcE homologs characterized to date are not identical. In particular, the *Vibrio* pathogenicity island-encoded virulence factor TagA, which shares 37% sequence identity with StcE, cleaves CD43 less efficiently compared to StcE and is inactive toward C1-INH (Szabady et al., 2011). This suggests that sequence divergence between these proteins may be responsible for differences in substrate preference. We hypothesize that a dynamic, sequence-variable region (INS, His132-Asn251) within the N-terminal IG domain, which likely has arisen from a gene insertion event, contributes to substrate specificity. The INS domain structure, which we solved separately, reveals complimentary surface features that are oriented toward the active site, indicating a possible exosite role for specifying binding partners as in the metalloprotease RVV-X (Takeda et al., 2007) (Figures 5A–5C). This is supported by our observed decrease in cleavage efficiency of ΔINS toward C1-INH. We speculate that StcE has evolved this modular architecture to fine-tune its substrate binding. Though present in most StcE-like proteins in the M66 family, select members lack this insertion element (for instance, EEZ85628 and EDL70231 from the marine bacteria *Vibrio harveyi* 1DA3 and HYϕ1, respectively) and instead, feature tandem chitin/cellulose-binding domains in the C termini as annotated in the Entrez protein sequence database (Figure S2). The variation in domain composition and organization likely reflects adaptation to changing substrate motifs in different host environments (Pruzzo et al., 2008). Truncation mutants and chimeric proteins designed based on the StcE model will provide insight in that regard.

The StcE structure reveals how the M66 peptidase family can acquire noncatalytic domains potentially conferring distinct substrate specificities while preserving essential features of its catalytic scaffold. The overall conformation of StcE and local flexibility in the active site appear to have evolved to specifically recognize O-glycan induced conformations in mucin-type glycoproteins. The findings provide a structural basis for understanding how bacterial virulence factors can breach the mucosal barrier and disarm key host immune proteins and specifically represent a highly attractive, accessible antimicrobial target to block enterohemorrhagic virulence and disease. Owing to its ability to cleave select mucins, StcE has also been proposed as a treatment for cystic fibrosis and shown to loosen the mucus in affected patients (US patent no. 7214479). The model presented here may also facilitate the design of new therapeutics for this devastating respiratory disease.

EXPERIMENTAL PROCEDURES

Cloning

The gene encoding *Escherichia coli* O157:H7 *stcE* residues 36–989 (accession code: O82882) was amplified from genomic DNA using the primers 5′-CCG CAGGCTAGCGCTGATAATAATTCAGCC-3′ and 5′-CCGCAGCTCGAGTTATT TATATACAACCCTCATTGACC-3′. The PCR product was ligated into NheI and XhoI sites of pET28b expression vector (Novagen). The pET28b*stc*EΔ35 construct served as a template for all subsequent StcE variants. 5′-GGGAATG AGTTCAGTCATGACGTTGGTCATAATTATGG-3′ and 5′-CCATAATTATGACC AACGTCATGACTGAACTCATTCCC-3′ were used to introduce the E447D mutation by QuickChange (Stratagene). The surface entropy reduction mutant K318A/K320A/E321A was generated by the same procedure with 5′-CGGG ATCGCTTTGATTTTGCCGCAGACGCAGCAGCACATAGGG-3′ and 5′-CCCT ATGTGCTGCTGCGTCTGCGGCAAAATCAAAGCGATCCCG-3′. The insertion deletion (ΔG140-F243) was chosen to be consistent with the StcE model and secondary structure prediction and removed by QuickChange using 5′-GGATGGTGTTCCGGAAGGTCGCTCCGGTGAACTGGAG and 5′-CTCCA GTTCACCGGAGCGACCTTCCGGAACACCATCC. Domain INS (H132-N251) was cloned using 5′-CCGCAGGCTAGCCATCTGGATGGTGTTCCGGAAG-3′

and 5′- CCGCAGCTCGAGTTAATTATTCTCCAGTTCACCGGAG-3′. All constructs were confirmed by sequencing.

Protein Expression and Purification

E. coli BL21 (λDE3) transformed with *stcE* variants was grown at 37°C until midexponential phase and protein production was induced with 0.3 mM IPTG at 20°C overnight. Harvested cells were disrupted by a pressurized homogenizer (Avestin) in lysis buffer (20 mM HEPES [pH 7.5] and 500 mM NaCl). The clarified lysate containing the target protein was purified sequentially by zinc-chelating Sepharose, MonoQ anion-exchange and Superdex-200 HR 10/30 columns (GE Healthcare) to >95% purity as judged by SDS-PAGE. The affinity tag was cleaved with thrombin prior to size exclusion and the protein stored in 20 mM HEPES (pH 7.5) and 150 mM NaCl.

Crystallization and Data Collection

Initial needle clusters were obtained from StcEΔ35E447D. After various attempts to improve the diffraction quality, larger crystals of the quadruple mutant K318A/K320A/E321A/E447D were grown by the hanging-drop method at 18°C. The drops consisted of a 1:1 ratio of StcE (7 $mg.ml^{-1}$) and the reservoir solution (0.2 M $MgCl_2$, 8% PEG8000, and 0.1 M MES [pH 6.5]). Crystals of domain INS were obtained with 26% PEG2000MME and 0.1M trisodium citrate (pH 5.5) and derivatized with 0.5 M NaI for halide phasing. Single crystals were cryoprotected by briefly soaking in the mother liquor supplemented with 25% glycerol and flash-frozen in liquid nitrogen. Since most of the quadruple mutant crystals exhibited anisotropy, extensive screening was done at the Canadian Light Source 08ID-1 beamline before sufficient diffraction was obtained from a SeMet derivative. Data were collected at the selenium peak (0.97907 Å) and processed using MOSFLM (Battye et al., 2011) and SCALA (Collaborative Computational Project, Number 4, 1994).

Structure Determination and Refinement

Initial experimental phases were obtained by SHARP of the automatic structure solution program autoSHARP (Vonrhein et al., 2007), which located 10 out of a possible 17 selenium sites. Subsequent solvent flipping with SOLOMON from autoSHARP resulted in a reasonably interpretable map for the core of StcE. The model was built manually using Coot (Emsley et al., 2010). Iterative rounds of refinement were performed using REFMAC (Murshudov et al., 1997) with TLS (translation/libration/screw) (Painter and Merritt, 2006), and BUSTER (Blanc et al., 2004) to improve modeling of the poorly resolved regions. The TLS parameters were generated by the TLS motion determination server, which modeled the flexible StcE structure as contiguous groups of atoms undergoing rigid-body motion. The optimal TLS groups, as judged by the diminishing net TLS residual, are composed of residues 39–115, 116–139 and 248–310, 311–411, 412–512, 513–519, 520–685, 686–800, and 801–898. The high Wilson B (61 $Å^2$) and overall refined B (83 $Å^2$) are perhaps not surprising, given the modular and dynamic nature of StcE. Asn856, which is located in a turn in domain C, lies in a disallowed region of the Ramachandran plot in addition to Ala41 in the poorly resolved N terminus. Initial iodide phases for domain INS were also obtained using autoSHARP. Following similar procedures for model building and refinement, its orientation in the overall StcE structure was determined by spherically averaged phased translation using SeMet phases from the quadruple mutant. We compared the relative orientation between M-SD1 and M-SD2 among StcE homologs by measuring the angle between structurally equivalent residues (Cα of 3UJZ, R372/H446/P535; 1IAG, S71/H142/D180; 2ERO, T267/H335/D373; 2DW2, A263/H333/N371). Structure figures were prepared with PyMol (http://pymol.sourceforge.net). The ConSurf server was used to identify functionally and structurally important residues (Glaser et al., 2003; Landau et al., 2005).

Static Light Scattering

Protein samples (0.5–1 $mg.ml^{-1}$) were injected onto a Superdex-200 column connected in line to a miniDAWN multiangle light scattering machine with an interferometric refractometer (Wyatt Technologies). The associated ASTRA software was used to estimate molecular mass for the data analysis.

Peptide docking

Molecular docking was performed with AutoDock Vina, using default values for the docking parameter file (Trott and Olson, 2010). The program merged all nonpolar hydrogens and assigned charges to the atoms in the coordinate files. A grid box of 22 × 22 × 18 grid points of 1 Å spacing covering the immediate active site region was defined as the search space. Docking results were manually inspected for structural and chemical complementarities. One of the evaluation criteria was that the distance between zinc and the P1 carbonyl oxygen be less than 3 Å.

Proteolytic Assay

StcE variants were incubated with C1-INH (Molecular Innovations) in a 1:20 (wt/wt) ratio in 20 mM HEPES (pH 7.5) and 150 mM NaCl at 37°C. Ten microliter aliquots were removed from the 100 μl total assay volume at 0, 15, 45, 150, 240, and 360 min, 23 hr, and 48 hr, and the reaction was terminated with EDTA (final concentration 50 mM). Bands corresponding to C1-INH were visualized using Pro-Q Emerald 300 glycoprotein stain (Molecular Probes).

Sucrose Density Gradient Centrifugation

Pellets from 200 ml cultures of *E. coli* BL21 (λDE3) expressing StcE variants were lysed by a pressurized homogenizer in buffer A (20 mM Tris [pH 7.5]). Cell debris was removed by a low-speed spin at 10,000 × *g* (JA10 rotor, Beckman), followed by a high-speed spin at 140,000 × *g* (Ti60 rotor, Beckman) to isolate the total membranes. The membrane pellet was washed in buffer A, centrifuged again and homogenized in 2 ml buffer A. Membrane fragments were then layered onto the sucrose gradient consisting of 0.8 ml 55%, 2 ml 50%, 2 ml 45%, 2 ml 35%, and 0.8 ml 30% and centrifuged in a SW41 swinging-bucket rotor (Beckman) at 180,000 × *g* for 16 hr.

Circular Dichroism Spectroscopy

CD spectra of 0.25 $mg.ml^{-1}$ WT StcE and mutant devoid of the insertion element were recorded in a Jasco J-810 spectropolarimeter at 25°C. All spectra were the average of four scans. The optical path length is 0.1 cm.

Glycan Array Screening

The glycan array-binding analysis was performed at the Consortium for Functional Glycomics (http://www.functionalglycomics.org/static/index.shtml). StcEΔ35E447D, fluorescently labeled with the Alexa Fluor 488 dye (Molecular Probes), was screened with version 4.2 of the printed mammalian glycan microarray containing 511 ligands in replicates of 6. The highest and lowest fluorescence readings were excluded from statistical analysis of the data in an attempt to eliminate false hits.

ACCESSION NUMBERS

The structure factors and coordinates have been deposited in the Protein Data Bank under accession codes 3UJZ and 4DNY.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and can be found with this article online at doi:10.1016/j.str.2012.02.015.

ACKNOWLEDGMENTS

We thank Dr. Gerd Prehna for the ZirT control sample, Dr. Thomas Spreter for help with light scattering analysis and Dr. Michael Gretes, Dr. Andrew Lovering and Ms. Liza de Castro for contributions at earlier stages of the project. We also thank the staff at the CMCF beamlines at the Canadian Light Source (Saskatoon, SK) for data collection. Services provided by The Consortium for Functional Glycomics were funded by the NIGMS GM62116. This work was funded by operating grants from the Canadian Institutes of Health Research and the Howard Hughes Medical Institute International Scholar Program, with infrastructure funding from the Michael Smith Foundation for Health Research, the Canada Foundation for Innovation, and the B.C. Knowledge Development Fund. A.C.Y.Y. was supported by the Michael Smith Foundation for Health Research. N.C.J.S is supported by the Canadian Research Chair Tier 1 program.

Received: November 30, 2011
Revised: February 11, 2012
Accepted: February 16, 2012
Published: April 3, 2012

REFERENCES

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. D Biol. Crystallogr. *67*, 271–281.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr. D Biol. Crystallogr. *60*, 2210–2221.

Bock, S.C., Skriver, K., Nielsen, E., Thøgersen, H.C., Wiman, B., Donaldson, V.H., Eddy, R.L., Marrinan, J., Radziejewska, E., Huber, R., et al. (1986). Human C1 inhibitor: primary structure, cDNA cloning, and chromosomal localization. Biochemistry *25*, 4292–4301.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. *36* (Web Server issue), W197–W201.

Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Coltart, D.M., Royyuru, A.K., Williams, L.J., Glunz, P.W., Sames, D., Kuduk, S.D., Schwarz, J.B., Chen, X.T., Danishefsky, S.J., and Live, D.H. (2002). Principles of mucin architecture: structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glycodomains. J. Am. Chem. Soc. *124*, 9833–9844.

Cyster, J.G., Shotton, D.M., and Williams, A.F. (1991). The dimensions of the T lymphocyte glycoprotein leukosialin and identification of linear protein epitopes that can be modified by glycosylation. EMBO J. *10*, 893–902.

Derrien, M., van Passel, M.W., van de Bovenkamp, J.H., Schipper, R.G., de Vos, W.M., and Dekker, J. (2010). Mucin-bacterial interactions in the human oral cavity and digestive tract. Gut Microbes *1*, 254–268.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Gal-Mor, O., Gibson, D.L., Baluta, D., Vallance, B.A., and Finlay, B.B. (2008). A novel secretion pathway of Salmonella enterica acts as an antivirulence modulator during salmonellosis. PLoS Pathog. *4*, e1000036.

Garner, B., Merry, A.H., Royle, L., Harvey, D.J., Rudd, P.M., and Thillet, J. (2001). Structural elucidation of the N- and O-glycans of human apolipoprotein(a): role of o-glycans in conferring protease resistance. J. Biol. Chem. *276*, 22200–22208.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163–164.

Gomis-Rüth, F.X. (2003). Structural aspects of the metzincin clan of metalloendopeptidases. Mol. Biotechnol. *24*, 157–202.

Gomis-Rüth, F.X. (2009). Catalytic domain architecture of metzincin metalloproteases. J. Biol. Chem. *284*, 15353–15357.

Grams, F., Dive, V., Yiotakis, A., Yiallouros, I., Vassiliou, S., Zwilling, R., Bode, W., and Stöcker, W. (1996). Structure of astacin with a transition-state analogue inhibitor. Nat. Struct. Biol. *3*, 671–675.

Grys, T.E., Siegel, M.B., Lathem, W.W., and Welch, R.A. (2005). The StcE protease contributes to intimate adherence of enterohemorrhagic Escherichia coli O157:H7 to host cells. Infect. Immun. *73*, 1295–1303.

Hang, H.C., and Bertozzi, C.R. (2005). The chemistry and biology of mucin-type O-linked glycosylation. Bioorg. Med. Chem. *13*, 5021–5034.

Hashimoto, R., Fujitani, N., Takegawa, Y., Kurogochi, M., Matsushita, T., Naruchi, K., Ohyabu, N., Hinou, H., Gao, X.D., Manri, N., et al. (2011). An efficient approach for the characterization of mucin-type glycopeptides: the effect of O-glycosylation on the conformation of synthetic mucin peptides. Chemistry *17*, 2393–2404.

Ho, T.D., Davis, B.M., Ritchie, J.M., and Waldor, M.K. (2008). Type 2 secretion promotes enterohemorrhagic Escherichia coli adherence and intestinal colonization. Infect. Immun. *76*, 1858–1865.

Holland, D.R., Tronrud, D.E., Pley, H.W., Flaherty, K.M., Stark, W., Jansonius, J.N., McKay, D.B., and Matthews, B.W. (1992). Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. Biochemistry *31*, 11310–11316.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38* (Web Server issue), W545–W549.

Holmes, N. (2006). CD45: all is not yet crystal clear. Immunology *117*, 145–155.

Inouye, M., Inouye, S., and Zusman, D.R. (1979). Biosynthesis and self-assembly of protein S, a development-specific protein of Myxococcus xanthus. Proc. Natl. Acad. Sci. USA *76*, 209–213.

Jentoft, N. (1990). Why are proteins O-glycosylated? Trends Biochem. Sci. *15*, 291–294.

Kurisu, G., Kinoshita, T., Sugimoto, A., Nagara, A., Kai, Y., Kasai, N., and Harada, S. (1997). Structure of the zinc endoprotease from Streptomyces caespitosus. J. Biochem. *121*, 304–308.

Lambris, J.D., Ricklin, D., and Geisbrecht, B.V. (2008). Complement evasion by human pathogens. Nat. Rev. Microbiol. *6*, 132–142.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. *33* (Web Server issue), W299–W302.

Lathem, W.W., Grys, T.E., Witowski, S.E., Torres, A.G., Kaper, J.B., Tarr, P.I., and Welch, R.A. (2002). StcE, a metalloprotease secreted by Escherichia coli O157:H7, specifically cleaves C1 esterase inhibitor. Mol. Microbiol. *45*, 277–288.

Lathem, W.W., Bergsbaken, T., and Welch, R.A. (2004). Potentiation of C1 esterase inhibitor by StcE, a metalloprotease secreted by Escherichia coli O157:H7. J. Exp. Med. *199*, 1077–1087.

Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V., and McGuckin, M.A. (2008). Mucins in the mucosal barrier to infection. Mucosal Immunol. *1*, 183–197.

Luo, Y., Li, S.C., Chou, M.Y., Li, Y.T., and Luo, M. (1998). The crystal structure of an intramolecular trans-sialidase with a NeuAc alpha2—>3Gal specificity. Structure *6*, 521–530.

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.

Nataro, J.P., and Kaper, J.B. (1998). Diarrheagenic Escherichia coli. Clin. Microbiol. Rev. *11*, 142–201.

Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D Biol. Crystallogr. *62*, 439–450.

Paton, A.W., and Paton, J.C. (2002). Reactivity of convalescent-phase hemolytic-uremic syndrome patient sera with the megaplasmid-encoded TagA protein of Shiga toxigenic Escherichia coli O157. J. Clin. Microbiol. *40*, 1395–1399.

Perkins, S.J., Smith, K.F., Amatayakul, S., Ashford, D., Rademacher, T.W., Dwek, R.A., Lachmann, P.J., and Harrison, R.A. (1990). Two-domain structure of the native and reactive centre cleaved forms of C1 inhibitor of human complement by neutron scattering. J. Mol. Biol. *214*, 751–763.

Pillai, L., Sha, J., Erova, T.E., Fadl, A.A., Khajanchi, B.K., and Chopra, A.K. (2006). Molecular and functional characterization of a ToxR-regulated lipoprotein from a clinical isolate of Aeromonas hydrophila. Infect. Immun. *74*, 3742–3755.

Pruzzo, C., Vezzulli, L., and Colwell, R.R. (2008). Global impact of Vibrio cholerae interactions with chitin. Environ. Microbiol. *10*, 1400–1410.

Rawlings, N.D., Barrett, A.J., and Bateman, A. (2010). MEROPS: the peptidase database. Nucleic Acids Res. *38* (Database issue), D227–D233.

 ©2012 Elsevier Ltd All rights reserved

Schoenberger, O.L. (1992). Characterization of carbohydrate chains of C1-inhibitor and of desialylated C1-inhibitor. FEBS Lett. *314*, 430–434.

Szabady, R.L., Lokuta, M.A., Walters, K.B., Huttenlocher, A., and Welch, R.A. (2009). Modulation of neutrophil function by a secreted mucinase of Escherichia coli O157:H7. PLoS Pathog. *5*, e1000320.

Szabady, R.L., Yanta, J.H., Halladin, D.K., Schofield, M.J., and Welch, R.A. (2011). TagA is a secreted protease of Vibrio cholerae that specifically cleaves mucin glycoproteins. Microbiology *157*, 516–525.

Takeda, S., Igarashi, T., and Mori, H. (2007). Crystal structure of RVV-X: an example of evolutionary gain of specificity by ADAM proteinases. FEBS Lett. *581*, 5859–5864.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. *31*, 455–461.

Van den Steen, P., Rudd, P.M., Dwek, R.A., and Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. Crit. Rev. Biochem. Mol. Biol. *33*, 151–208.

Veltman, O.R., Eijsink, V.G., Vriend, G., de Kreij, A., Venema, G., and Van den Burg, B. (1998). Probing catalytic hinge bending motions in thermolysin-like proteases by glycine —> alanine mutations. Biochemistry *37*, 5305–5311.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol. Biol. *364*, 215–230.

 ©2012 Elsevier Ltd All rights reserved