Molecular Recognition of Chymotrypsin by the Serine Protease Inhibitor Ecotin from Yersinia pestis

Elizabeth A. Clark†, Nicola Walker‡, Donna C. Ford‡, Ian A. Cooper‡, Petra C. F. Oyston‡ and K. Ravi Acharya†,1

From the †Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom and ‡Biomedical Sciences, Defence Science and Technology Laboratory, Porton Down, Salisbury SP4 0JQ, United Kingdom

1Address correspondence to: K. Ravi Acharya, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom; Tel: 44-1225-386238; E-mail: K.R.Acharya@bath.ac.uk

Running head: Crystal structure of Yersinia pestis ecotin in complex with chymotrypsin

*Supported by the Wellcome Trust (United Kingdom) through an equipment grant (number 088464) and a Royal Society Industry Fellowship to K.R.A.

Keywords: molecular recognition, ecotin, chymotrypsin, crystal structure, protein-protein interaction

Resistance to antibiotics is a problem not only in terms of healthcare but also biodefence. Engineering of resistance into a human pathogen could create an untreatable bioterror pathogen. One such pathogen is Yersinia pestis, the causative agent of plague. Previously we have used a bioinformatic approach to identify proteins which may be suitable targets for antimicrobial therapy and in particular for the treatment of plague. The serine protease inhibitor ecotin was identified as one such target. We have carried out mutational analyses in the closely related Yersinia pseudotuberculosis which has validated that the ecotin gene is a virulence-associated gene in this bacterium. Yersinia pestis ecotin inhibits chymotrypsin. Here we present the structure of ecotin in complex with chymotrypsin to 2.74 Å resolution. The structure features a biologically relevant tetramer whereby an ecotin dimer binds to two chymotrypsin molecules similar to what was observed in related serine protease-inhibitor structures. However, the vast majority of the interactions in the present structure are distinctive, indicating that the broad specificity of the inhibitor for these proteases is based largely on its capacity to recognize features unique to each of them. These findings will have implications for the development of small ecotin inhibitors for therapeutic use.

Antibiotic resistance is an ever-increasing problem in the fight against microbial disease. It is a major public health issue, where naturally resistant strains are emerging against all classes of antibiotics. In addition, genetic engineering has raised the spectre that resistance could be engineered into an organism in order to create an untreatable bioterror pathogen. Current bactericidal antibiotics inhibit only a few targets: DNA, RNA, cell wall or protein.
environmental resistance. Yersinia and the Madagascar pesta\natives are selenium pathogens, while Yersinia pestis is the aetiological agent of bubonic and pneumatic plague. Y. pestis is considered to be a clonal derivative of Y. pseudotuberculosis (2) but is awarded species designation due to the severity of the disease it causes. The infectivity of Y. pestis by the aerosol route, the acute disease and high associated mortality rate meant that plague attracted attention as a potential bioweapon in the offensive biological weapons programmes of the last century, and is still an agent of concern in biodefence. Antibiotic resistant strains are rare, but a multiple antibiotic-resistant strain was isolated in Madagascar (3) and efficient transfer of resistance genes to Y. pestis in the midgut of the flea has been demonstrated (4). Therefore, there is a need for novel effective therapeutics for use in treatment of plague.

It is notable that the close relatives Y. pestis and Y. pseudotuberculosis cause remarkably different diseases, yet are very closely genetically related (99% nucleotide identity for most shared genes), and data based on multiple locus sequence typing (MLST) analysis suggests that Y. pestis diverged from Y. pseudotuberculosis only 1,500-20,000 years ago (2). The evolution from a mild gut pathogen into a systemic pathogen, which has repeatedly devasted the human race, has occurred in an eye-blink of evolutionary time. A major step along this pathway appears to be the acquisition of two plasmids by Y. pestis. Pathogenic Y. pseudotuberculosis retains a single plasmid (pCD1), whereas Y. pestis also possesses the plasmids pMT1 and pPCP1. It is therefore widely accepted that Y. pestis was once a simple enteropathogen, and gene acquisition and loss allowed it to change its environmental niches and its lifestyle. The complete genome sequence of Y. pestis CO92 (biovar Orientalis) was determined and revealed a highly fluid genome, with large inversions occurring between IS elements (5). A large number of pseudogenes are also present in the Y. pestis genome, disrupted by IS elements, or by other mutations. These pseudogenes are thought to have been required for an enteropathogenic lifestyle, but are now redundant. However, the close relationship between Y. pestis and Y. pseudotuberculosis means that often the role of proteins in one organism retain the same role in the other. This allows us to gain insight into Yersinia virulence using the less pathogenic enteropathogen, which can then be followed up by focused efforts on the highly virulent plague bacillus.

When considering novel antimicrobial development, ideally a target will be common to a range of pathogens. We have previously reported a bioinformatic approach to identify genes more commonly found in pathogens than non-pathogens (6). One of the targets identified as potentially of interest was ecotin. Ecotin is a serine protease inhibitor that inhibits a range of serine proteases including trypsin, chymotrypsin and elastase (7). In the present study, we have validated ecotin as an attenuating locus. We have elucidated the crystal structure of ecotin in complex with activated bovine chymotrypsin and described the molecular basis for the serine protease inhibition by ecotin. These molecular details provide a useful platform on which to base further research into the development of novel therapeutics for plague.

**EXPERIMENTAL PROCEDURES**

Unless otherwise stated chemicals were purchased from Sigma-Aldrich (Poole, United Kingdom) and enzymes were purchased from Promega Ltd (Southampton, United Kingdom).
Preparation of *Y. pseudotuberculosis* mutants—Construction of *Y. pseudotuberculosis* mutants was carried out as reported previously (8). Briefly, primers were designed for the target gene to be disrupted that included 20 bp complementary to the 5′ or 3′ sequence of the kanamycin gene of the plasmids pK2 or pUC4K followed by 50 bp of upstream or downstream sequence flanking the gene to be disrupted. The following primers were used:

5′-AATCGAGTTTTAGAGCTATTGCTGATAAAAACCTA
GAGAAAAAAGATG-3′ (forward) and
5′-ATGTTAATATTAGCCAGCGCAGATGCCT
GGTTAATGATGCTA-3′ (reverse). PCR products were generated using the plasmid pK2 as a template and excess template was digested with *DpnI*. PCR products were purified using Millipore Microcon Ultragel YM-100 and were then transformed into *Y. pseudotuberculosis* YPIII pAJD434 (9) by electroporation. Following overnight incubation at 28 °C in Luria-Bertani (LB) supplemented with 0.8% arabinose, transformants were selected on LB agar supplemented with kanamycin (50 μg/ml) and trimethoprim (100 μg/ml) for 48 h at 28 °C. Transformants were screened by PCR using target gene-specific and kanamycin gene-specific primers; 5′-GATCCTACCTGTTGTTGCCT
3′ (forward) and 5′-TTAGCCAGCGGCGAATAG-3′ (reverse).

Mutant strains were cured of the pAJD434 plasmid by growth at 37 °C in LB broth media supplemented with kanamycin (50 μg/ml). Cured mutant strains were screened for the virulence plasmid pYV by PCR for two genes located on this plasmid; virF and yscC. The retention of the *Yersinia* virulence plasmid (pYV) was also confirmed by culture on Congo-red magnesium oxalate (CR-MOX) plates, where plasmid retention results in small red colonies and plasmid loss results in large pink colonies (10).

Competitive index studies—For *in vivo* competitive index studies, mutant and wild type strains were grown separately to exponential phase in 20 ml LB broth with shaking. Broth cultures were then centrifuged (10 minutes, 4000 g) and the pellet re-suspended in 10 ml sterile PBS and centrifuged again (10 minutes, 4000 g). The bacteria were washed and re-suspended in 10 ml PBS and the OD600 adjusted to 0.55 - 0.60 with sterile PBS. Wild type and mutant bacterial suspensions were then mixed in a 1:1 ratio and serially diluted with sterile PBS to give an inoculation concentration of approximately 1 × 10^3 cfu/ml. Groups of 6 mice were then dosed with 0.1 ml of this solution by the intravenous (i.v.) route. Retrospective viable counts were determined by plating out dilutions (in triplicate) on LB agar and LB agar supplemented with kanamycin to determine the input ratio. After 5 days, spleens were recovered and passed through 70 μm sieves (Becton Dickinson) to produce a cell suspension in 3 ml of PBS. Cell suspensions were serially diluted in sterile PBS and plated onto LB agar and LB agar supplemented with kanamycin to determine the output ratio. The competitive index (CI) is defined as the output ratio (mutant/wild type) divided by the input ratio (mutant/wild type) (11,12).

Cloning of the ecotin gene—The eco gene minus the signal peptide was PCR amplified from *Y. pseudotuberculosis* IP32953 DNA using the primers 5′-ACGCCTACGCTCTCAATCA-3′ (forward) and 5′-CTATGGACCGCGACTTCTGA-3′ (reverse) and was cloned into the expression vector pCRT/NT-TOPO (Invitrogen). The plasmid was transformed, by heat shock, into *E.coli* strain TOP 10F’. Resultant constructs were sequenced and those in which ecotin was appropriately orientated for translation were chosen and transformed into *E. coli* BL21*(DE3) cells for protein expression studies.
Expression and purification of ecotin—

*E. coli* BL21* (DE3) harboring recombinant pCRT7/NT-TOPO plasmids were cultured in LB broth supplemented with 1% glucose and ampicillin (50 μg/ml). Cultures were grown with shaking (170 rpm) at 37 °C until an OD$_{600}$ of 0.4 was reached. Protein expression was induced with 1 mM IPTG (isopropyl-B-D-thiogalactoside) with incubation for a further 4 hours followed by harvesting by centrifugation (10 minutes, 1700 g). Cell pellets were resuspended in PBS and were sonicated in an ice water bath. The suspension was clarified by centrifugation at 27000 g for 30 minutes. Supernatants were loaded onto a HisTrap™ column (GE Healthcare) equilibrated with 40 mM Tris pH 7.5, 750 mM NaCl. The column was washed with equilibration buffer supplemented with 10 mM and 20 mM imidazole prior to elution with elution buffer (40 mM Tris pH 7.5, 750 mM NaCl, 250 mM imidazole). Elution fractions containing ecotin as determined by SDS-PAGE analysis were pooled and buffer exchanged into 40 mM Tris pH 7.5.

Chymotrypsin inhibition assay—Inhibition of chymotrypsin activity was determined by hydrolysis of N-Benzoyl-L-Tyrosine Ethyl Ester (BTEE) based on a method by Wirnt (13). Varying concentrations of ecotin were mixed with 1.4 nM chymotrypsin (Sigma-Aldrich) in a 1:1 v/v ratio and incubated for 5 minutes at 25 °C prior to addition to a 1 ml reaction volume containing 465 nM BTEE in 80 mM Tris, 20 mM CaCl$_2$ pH 7.8 pre-warmed to the same temperature. The OD$_{256}$ was monitored for 3 minutes using an Ultraspec 4000 (GE Healthcare).

Crystallisation of ecotin-chymotrypsin complex—Ecotin at 3 mg/ml and chymotrypsin (Sigma-Aldrich) at 6 mg/ml in 40 mM Tris pH 7.5 were mixed in a 1:1 volume ratio before incubation at 16 °C for 2 hours prior to crystallisation trials. Initial screens to identify conditions in which the ecotin-chymotrypsin complex crystallized were carried out using pre-prepared 96-well screens (Molecular Dimensions) on an Art Robbins Phoenix nano dispenser. 300 nl drops were set-up using the sitting drop method. Crystals were obtained under various conditions and optimisation was carried out in 24-well plates using the hanging drop method. Improved crystals of the ecotin-chymotrypsin complex grew over ~ 8 weeks in a 2 μl drop which contained a 1:1 volume ratio of protein solution to mother liquor. Mother liquor contained 100 mM Bis Tris pH 5.5, 200 mM ammonium sulfate and 20 % (w/v) PEG 3350.

Structure determination and refinement—X-ray data were collected from a single cryocooled crystal with 25% PEG 3350 at the Diamond Light Source (Didcot, Oxon, UK) on an ADSC Q315 CCD detector on station I03 (λ=0.97 Å). Diffraction images were collected at an oscillation angle of 1.0°. Data were processed in the monoclinic space group $P2_1$ using the HKL2000 package (14). Initial phases were obtained through molecular replacement using the program PHASER (15). The search models used for molecular replacement were the 1.68 Å structure of bovine chymotrypsin (PDB code 4CHA) and a model of ecotin generated based on its sequence alignment with *E. coli* ecotin (PDB code 1ECZ). Model building was performed with Coot (16) and refinement was performed with REFMAC5 (17). A set of reflections was set aside for $R_{free}$ calculation (18).

RESULTS

Competitive index (CI) study—The ecotin gene was previously identified as being present predominantly in pathogenic bacteria whilst rarely found in non-pathogens (6). In order to validate ecotin as an attenuating locus when mutated in *Yersinia*, the gene encoding ecotin was disrupted in *Y. pseudotuberculosis*
to create a mutant. Mice were then infected with mutant and wild-type *Y.
psuedotuberculosis* (see Experimental procedures) and the CI was calculated. A CI
value of 0.2 or less indicates that the locus is attenuating. The CI of the ecotin-defective
mutant was 0.03 ± 0.02 (mean ± SD), indicating that this is a highly attenuating locus, thus
identifying ecotin as a virulence-associated protein.

*Ecotin-chymotrypsin inhibition assay*—An assay was carried out in order to determine
the effect of *Yersinia* ecotin on chymotrypsin. Chymotrypsin activity was tested subsequent
to incubation with varying concentrations of ecotin. The result, shown in figure 1, indicates
that chymotrypsin activity is impaired upon incubation with ecotin.

**The ecotin-chymotrypsin structure**

*Quality of the structure*—The crystallographic asymmetric unit contains four
copies of each of ecotin and chymotrypsin molecules (Figure 2) with a solvent content of
44 %. Molecular replacement with the aid of the structure of bovine chymotrypsin (PDB
code 4CHA) as a search model was used to place four molecules of chymotrypsin in the
asymmetric unit. The model of *Yersinia* ecotin based on its sequence alignment with *E.
coli* ecotin (PDB code 1ECZ) was then used to find two molecules of ecotin in the asymmetric
unit. Positioning of the third molecule of ecotin was carried out by rotation of a symmetry-
related molecule. The remaining difference density was indicative of the presence of a
fourth molecule of ecotin. However, the orientation was unclear. Model-building was
carried out, initially using poly-alanine chains, until the density was improved such that the
fourth molecule of ecotin could be positioned through rotation of a symmetry-related
molecule.

The final refined structure has an *R*\textsubscript{cryst}
of 24.4 % and an *R*\textsubscript{free} of 31.4 % (Table 1). The
final model consists of 33 water molecules and four sulfate ions in the asymmetric unit. The
Ramachandran plot generated by PROCHECK (19) shows that 88.5 % of residues are in
allowed regions, 11.4 % are in additionally allowed regions and 0.1 % (1 residue) is in a
generously allowed region.

Activated bovine chymotrypsin is
comprised of three disulphide-linked polypeptide chains of 11, 130 and 97 residues
and density can be observed for all three chains (Cys1 to Ser11, Ile16 to Tyr146 and
Ala149 to Asn245).

The functional *Yersinia* ecotin protein is
comprised of a single polypeptide chain (Asp22

---

The functional *Yersinia* ecotin protein is
comprised of a single polypeptide chain (Asp22
to Lys169). The electron density observed at
the N-terminus differs between each copy of
ecotin; commencing at Gln29, Asn28, Gln30
and Pro32 for chains E1, E2, E3 and E4,
respectively. Clear density was observed
to the C-terminal Lys169 residue in all
four copies of ecotin.

**Overall topology**—Chymotrypsin is a
globular, monomeric protein comprised of
discised twelve β-strands, two α-helices, and a number of loop regions. In the present complex
structure, chymotrypsin molecules

---

The functional *Yersinia* ecotin protein is
comprised of a single polypeptide chain (Asp22
to Lys169). The electron density observed at
the N-terminus differs between each copy of
ecotin; commencing at Gln29, Asn28, Gln30
and Pro32 for chains E1, E2, E3 and E4,
respectively. Clear density was observed
to the C-terminal Lys169 residue in all
four copies of ecotin.

**Overall topology**—Chymotrypsin is a
globular, monomeric protein comprised of
discised twelve β-strands, two α-helices, and a number of loop regions. In the present complex
structure, chymotrypsin molecules

---

The functional *Yersinia* ecotin protein is
comprised of a single polypeptide chain (Asp22
to Lys169). The electron density observed at
the N-terminus differs between each copy of
ecotin; commencing at Gln29, Asn28, Gln30
and Pro32 for chains E1, E2, E3 and E4,
respectively. Clear density was observed
to the C-terminal Lys169 residue in all
four copies of ecotin.

**Overall topology**—Chymotrypsin is a
globular, monomeric protein comprised of
discised twelve β-strands, two α-helices, and a number of loop regions. In the present complex
structure, chymotrypsin molecules

---

The functional *Yersinia* ecotin protein is
comprised of a single polypeptide chain (Asp22
to Lys169). The electron density observed at
the N-terminus differs between each copy of
ecotin; commencing at Gln29, Asn28, Gln30
and Pro32 for chains E1, E2, E3 and E4,
respectively. Clear density was observed
to the C-terminal Lys169 residue in all
four copies of ecotin.
the previously published structure of ecotin complex (PDB code 1N80) ranges from 0.82 to 0.92 Å. Ecotin takes the form of a homodimer that binds to two chymotrypsin molecules. The ecotin dimer is shaped like a ‘butterfly’ and this shape provides two clefts into which chymotrypsin can fit (Figure 2). Binding of two molecules of chymotrypsin to the ecotin dimer is facilitated through two distinct binding sites for chymotrypsin on monomeric ecotin. A primary interaction occurs between ecotin and the active site of chymotrypsin and a secondary interaction between the partner ecotin and a different part of chymotrypsin acts to stabilize binding. This results in a biologically relevant tetramer which in the structure forms half of the asymmetric unit (Figure 2). The shape complementarity between the dimer and each chymotrypsin monomer is good; with an average $S_c$ value of 0.73 where $S_c$ is the shape correlation statistic (20). Details of the interfaces found within the asymmetric unit due to molecular recognition of the complex are shown in Table 2. Details on buried surface areas and solvation free energy gain for the interfaces discussed below were calculated using the program PISA (21).

**The ecotin homodimer** — The C-terminal arm of ecotin - residues Glu158 to Lys169 - exchanges with its dimeric partner to create a two-stranded anti-parallel β-ribbon. 10 potential hydrogen bonds exist between each exchanged arm and residues of the partner ecotin molecule giving a total of 20 potential hydrogen bonds between monomers of the E1/E2 dimer. These hydrogen bonds are predominantly between main chain atoms of the exchanged arm and residues Val152 – Lys169 of the partner ecotin. 15 % of the solvent accessible area of each ecotin monomer is buried at the dimer interface giving a total buried surface area of 3142 Å².

**The primary ecotin-chymotrypsin binding interface (E1/C1)** — Two extended loops of ecotin bind at the chymotrypsin active site; an outer loop and a supporting inner loop. 10 % of the solvent-accessible area of ecotin is buried at the interface with chymotrypsin and the total buried surface area is 1924 Å² with a solvation free energy gain of -14.3 kcal/mol (21). There are some 25 ecotin residues at the interface. The interface residues of ecotin are predominantly derived from the two loop regions of ecotin - residues Gln105 – Glu115 on the outer loop and residues Asp75 – Asn84 on the inner loop. Additional contacts are made by Thr125 and Asn127. The extended loops are stabilized by a disulphide bridge between Cys76 on the inner loop and Cys113 on the outer loop. In addition there are 13 potential hydrogen bonds between each loop, predominantly between main chain atoms.

There are 35 chymotrypsin residues at the interface with ecotin. These residues are predominantly derived from loop regions at and adjacent to the active site which is comprised of a catalytic triad of residues His57, Ser195 and Asp102. There are 11 potential hydrogen bonds between ecotin and chymotrypsin at, and adjacent to, the chymotrypsin active site. Met110 of ecotin is known as the P₁ residue, based on the standard nomenclature of binding sites as described by Schechter and Berger (22). Met110 resides on the main binding loop of ecotin and is involved in three hydrogen bonds; the main chain oxygen atom of ecotin Met110 faces the oxyanion hole of chymotrypsin and is hydrogen bonded to the main chain nitrogen atoms of Ser195 and Gly193. There is also a hydrogen bond between the main chain nitrogen atom of Met110 and the OG oxygen of Ser195 and the Met110 side chain fits into the substrate binding pocket of chymotrypsin (Figure 3). In addition to hydrogen bonding, a number of van der Waals interactions occur (Table 2). Residues at the interface display chemical character as follows; 45 % hydrophobic, 40 % polar uncharged and 15 % charged.
The secondary ecotin-chymotrypsin binding interface (E1/C2) — The second binding site is located at the other end of the elongated ecotin molecule. There are 14 ecotin residues at this interface (Thr89 to Asp96 and Arg135 to Leu140). This second binding site serves to stabilize the first binding site with about 5% of the accessible surface area of ecotin being buried at this site to give a total buried surface area of 1051 Å². The solvation free energy gain upon formation of the interface ranges from -3.3 kcal/mol to -0.7 kcal/mol with an average of -1.8 kcal/mol. This low number reflects the fact that this binding site is subsidiary. Despite the subsidiary nature of this binding site, there are at least four hydrogen bonds between ecotin and chymotrypsin at this site along with a number of van der Waals interactions (Table 2). The large buried surface area and the presence of a number of van der Waals contacts and potential hydrogen bonds identify this as biologically relevant interface as opposed to a crystal contact interface. The most extensive interactions involve the hydrophobic ecotin residues Trp93 and Phe95 of the Thr89 - Asp96 loop and Leu140, Arg135 and Arg139 from the Arg135 - Leu140 loop (Figure 4). There are 18 interfacing chymotrypsin residues at the secondary binding site. These residues are predominantly derived from the 90s loop and the Thr232 - Gln240 section of the chymotrypsin C-terminal α-helix. Residues at the interface are 45% hydrophobic, 33% polar uncharged and 21% charged.

Variation between interfaces — Local variation between the four copies of the interfaces in terms of which atoms are involved in hydrogen bonding and van der Waals contacts occurs as a result of alternative arginine side chain conformations. For example, at the E1/C1 primary interface, there is hydrogen bond between the side chain of ecotin Arg78 and the side chain of Tyr94 of chymotrypsin. At the E2/C2 primary interface there is a potential hydrogen bond between ecotin Arg78 and Ser96 of chymotrypsin. At the E3/C3 and E4/C4 interfaces potential hydrogen bonds are not observed for this arginine side chain. At the E3/C4 and E4/C3 secondary interfaces potential hydrogen bonds are observed between side chain atoms of Arg135 and main chain atoms of Lys93 and Tyr94, interactions which are not observed in the E1/C2 and E2/C1 interfaces. Despite this local variation, key recognition interactions, such as those involving residues of the chymotrypsin catalytic triad, are conserved.

Ligands — Four sulfate molecules (from the crystallization medium) are bound in the complex structure. One sulfate molecule resides at the periphery of each secondary binding site. The area is solvent-exposed and the sulfate molecules are close to the side chains of chymotrypsin residues Asn95 and Lys93, and ecotin residue Arg135. All four sulfate molecules are within hydrogen bonding distance of the side chain of the chymotrypsin residue Asn95. Additional hydrogen bonding characteristics vary between the four sulfate molecules due to the variation in side chain conformation of adjacent residues - one sulfate molecule is within hydrogen bonding distance of the nearby chymotrypsin Lys93 side chain and two sulfate molecules are within hydrogen bonding distance of the nearby ecotin Arg135 side chain. Sulfate-mediated hydrogen bonding between ecotin and chymotrypsin provides additional stabilization at these interfaces (Figure 4B).

DISCUSSION

The need for new antibiotics has become pressing in light of the emergence of antibiotic resistant strains of human pathogens. Development of such drugs is not only important in terms of healthcare, but also in biodefence. Y. pestis, the causative agent of plague, is an agent of concern in biodefence. One potential new target for the development
of novel antimicrobials is the serine protease inhibitor ecotin. The gene encoding ecotin was found to be more common in the genomes of pathogens than non-pathogens therefore identifying ecotin as a potential target (6). We have confirmed that eco is an attenuating locus in Y. pseudotuberculosis, identifying it as a potential target in this bacterium and the closely related bacterium Y. pestis. It has previously been postulated that ecotin plays a role in pathogenicity in other pathogens; the presence of ecotin in the E. coli periplasm led to the suggestion that ecotin acts to protect the bacterium from serine proteases in the gut of the mammalian host (7). As Y. pseudotuberculosis is also an enteric pathogen it is possible that it plays a similar role in this organism. However, we observed attenuation after intravenous challenge, thus by-passing the gut, and the gene is functional in Y. pestis, implying further functions in pathogenesis. Yersinia ecotin inhibits neutrophil elastase and so there is a potential role for ecotin in protection against intracellular host immunity (23). Yersinia are susceptible to killing by neutrophils and ablation of neutrophils results in increased colonisation (24), therefore it is possible that ecotin’s main function in vivo is to protect Yersinia from neutrophil killing. It has been observed that E. coli ecotin enhances recovery following neutrophil elastase treatment (23).

Yersinia ecotin inhibits neutrophil elastase, trypsin, chymotrypsin, FXa and thrombin (23). This pan-specificity is shared by E. coli ecotin (25,26) and is thought to be partly facilitated by the presence of a methionine at the P₁ position. Serine-protease inhibitors traditionally have a target-specific residue at the P₁ position whereas a methionine is considered to be a good ‘all-rounder’. It has been suggested that the contribution of the secondary binding site to the free energy of the E. coli ecotin-trypsin complex (27) allows for the less than favorable methionine at the P₁ site (28).

We have observed inhibition of chymotrypsin by Yersinia ecotin and we have provided a structural basis for this observed inhibition. Ecotin inhibits chymotrypsin in a 'substrate-like' fashion (32) and the Kᵢ for Yersinia ecotin inhibition of chymotrypsin is <0.002 nM (23). The 100 % sequence identity of ecotin across a range of species of Yersinia makes this structural result relevant to a range of Yersinia species including Y. pestis, although inhibition of chymotrypsin may be less important for Y. pestis than for Y. pseudotuberculosis. We have shown that the Yersinia ecotin – chymotrypsin complex features a biologically relevant tetramer, in line with what is observed in a number of related structures. A number of structures of E. coli ecotin in complex with a target protease have identified the biologically significant unit as a tetramer which is formed when a dimer of ecotin binds two target proteases. In the structure of the of E. coli ecotin – rat ionic trypsin complex there is one tetramer per asymmetric unit (27). Wang et al., (29) created an E. coli ecotin mutant and solved the structure of the mutant in complex with thrombin and the asymmetric unit in this structure is comprised of half of the tetramer. The crystal packing in the Yersinia ecotin – chymotrypsin complex shown here has displayed the unique arrangement of the asymmetric unit which contains two tetramers related by non-crystallographic symmetry. A key element of the biologically relevant tetramer is the ecotin dimer. Dimerisation of ecotin provides a large surface area – 17802 Å² for the Yersinia ecotin-chymotrypsin complex - which features two clefts into which two molecules of chymotrypsin can fit. This large surface area has meant that E. coli ecotin has attracted attention as a potential scaffold for engineering novel protease inhibitors (30).

The primary interface between ecotin and chymotrypsin buries a large surface area. The buried surface area of 1927 Å² is comparable to that of the related structures of
E. coli ecotin – protease complexes; the primary interface of the ecotin-trypsin complex buries 2000 Å² (27), that of the ecotin mutant-thrombin complex buries 2290 Å² (29) and the ecotin-crab collagenase complex buries 1950 Å² at this interface (31). The ecotin binding loop on which the P₁ site methionine resides is stabilized through binding to an inner loop. There is a disulphide bridge between the two loops anchored by a number of hydrogen bonds. The architecture of this binding loop is characteristic for that of the ‘substrate-like’ family of serine protease inhibitors and their protease ligands (32). The hydrogen bonding characteristics of the P₁ methionine residue are also characteristic of this type of inhibitor.

Some specific comparisons can be made between the Yersinia ecotin - chymotrypsin primary interface and that of the related E. coli ecotin - protease interfaces. The Kᵢ value for E. coli ecotin is also <0.002nM (23) hence any structural differences between the E. coli and Yersinia pestis ecotins should not affect its ability to inhibit chymotrypsin. Yersinia ecotin and E. coli ecotin share 70 % identity and 79 % similarity across the C-terminal 139 residues of the functional protein (Figure 5). Key residues are conserved between E. coli and Yersinia ecotin residues at the primary interface including the methionine residue at the P₁ position. Hydrogen bonding features of the P₁ residue in E. coli ecotin at the interface with chymotrypsin (PDB code 1N8O) are also found in this structure and the related E. coli ecotin - rat anion trypsin structure (27). Amongst the residues of ecotin at binding site 1, Ser108 - Pro114 (P₃ - P₄) are conserved. Adjacent to this section is a glutamic acid residue (Glu115) not present in E. coli ecotin, the side chain of which extends towards chymotrypsin in the Yersinia structure and makes contacts. The P₆ serine of E. coli ecotin is substituted for a glutamine (Gln105) in Yersinia ecotin, which contributes more to the solvation free energy gain upon complex formation than the equivalent serine and makes a potential hydrogen bond with chymotrypsin. Arg78 on the inner loop of Yersinia ecotin shows potential hydrogen bonds in two of the four interfaces; it is replaced with a leucine residue in E. coli ecotin which is not involved in hydrogen bonding. Asn84 of Yersinia ecotin lies at the periphery of the residues involved in binding at binding site 1, extending this section in comparison to E. coli ecotin, where the substituted lysine makes no contacts with chymotrypsin.

Ecotin is unique amongst serine protease inhibitors in that it makes an additional discrete interaction with the protease. In the absence of structural data for the secondary ecotin – chymotrypsin interaction, McGrath et al. (27) modeled ten putative secondary site interactions for the E. coli ecotin – chymotrypsin complex, nine of which featured ecotin residues which are conserved between E. coli and Yersinia ecotin. Our structural data, however (based on the structure of the Yersinia ecotin-chymotrypsin complex) shows that only three of those hydrogen bonds are predicted correctly in the modeled structure.

Various features of the ecotin-chymotrypsin structure reflect the inherent flexibility of ecotin. It can be seen that the solvation free energy gain upon complex formation for the primary and secondary interfaces varies throughout the asymmetric unit. There is a complementation process whereby the sum of the energy gain for the interfaces within each tetramer in the asymmetric unit is approximately the same however the distribution of energies across the interfaces within the tetramers varies; compare the solvation free energy gain upon complexation at the E1/C1 interface of -14.3 kcal/mol and that of the E1/C2 interface of -3.3 kcal/mol, with the equivalent primary and secondary interface values of -16.7 kcal/mol (E3/C3) and -0.7 kcal/mol (E3/C4) in the E3/E4/C3/C4 tetramer. It can be postulated that alterations in binding occur as a result of crystal packing and that the flexibility of ecotin allows
the molecule to compensate for this by altering binding elsewhere, resulting in a stable tetrameric complex arrangement. Flexibility is a trait shared with E. coli ecotin which has been characterized as an inherently flexible molecule (28).

Conclusions

Ecotin has been identified as a potential protein target for antimicrobial development and we have validated the eco gene as an attenuating locus in Y. pseudotuberculosis, identifying that ecotin is a virulence factor for this bacterium. Ecotin contributes to Yersinia pathogenicity through its inhibition of host serine proteases and the ecotin – chymotrypsin structure described represents an important host – pathogen interaction. The molecular details shown here will provide an adequate platform on which to base further research; an inhibitor of the ecotin – protease interaction, for example, may reduce the pathogenicity of Y. pestis such that it has potential for development as a treatment for plague.

Acknowledgments - We thank the scientists at station IO3, Diamond Light Source, Didcot (Oxon, UK) and Nethaji Thiagararajan for their support during X-ray data collection. We would also like to thank Helen E. Jones, Carwyn Davies and Melanie L. Duffield for technical support.

The atomic coordinates and structure factors for the ecotin-chymotrypsin complex (with code 2y6t) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/)

REFERENCES

FIGURE LEGENDS

Figure 1. Chymotrypsin inhibition assay data. Chymotrypsin activity was assayed subsequent to incubation with varying concentrations of ecotin. 1.4 nM chymotrypsin was incubated with no ecotin (lines), 0.5 nM ecotin (squares), 2.5 nM ecotin (triangles) or 3.5 nM ecotin (crosses) prior to assaying. The OD$_{256}$ of the BTEE hydrolysis reaction mixture was measured at various time intervals over the 180 s time course. Reaction rate is related to change in OD$_{256}$ and is indicative of chymotrypsin activity.

Figure 2. The structure of ecotin-chymotrypsin complex. Chymotrypsin molecules are coloured green and are designated E1 – E4. Ecotin molecules are coloured pink, yellow, red and blue and are designated C1 – C4. (A) There are two tetramers in the asymmetric unit, orientated as shown. (B) Cartoon depiction of molecules in the asymmetric unit. Figure prepared using PyMOL (www.pymol.org).

Figure 3. The ecotin binding loop at the chymotrypsin interface. (A) Residues of the ecotin Gln105 – Pro114 binding loop are represented as sticks and are coloured pink. Oxygen, nitrogen and sulfur atoms are coloured red, blue and yellow, respectively. The ecotin Cys113 side chain has been omitted for clarity. A surface representation of chymotrypsin is shown and residues which make direct contacts with ecotin are coloured green. The side chain of Met110 can be seen extending into the chymotrypsin substrate binding pocket. (B) Hydrogen bonding of ecotin Met110 and Ala112 are represented by dashed lines and contacting chymotrypsin residues are shown as green sticks. Figure prepared using PyMOL (www.pymol.org).

Figure 4. Interactions at the secondary ecotin – chymotrypsin interface (E3 – C4). Ecotin is coloured pink and chymotrypsin is coloured green. Interfacing residues from each protein are shown as sticks and oxygen and nitrogen atoms are coloured red and blue, respectively. Atoms that are within hydrogen bonding distance are connected by black dashed lines and van der Waals interactions are indicated by orange dashed lines. (A) Hydrogen bonding and van der Waals interactions between residues on the Thr89 – Asp96 loop and residues of the chymotrypsin C-terminal α-helix are shown. The potential hydrogen bond between Lys93 and Gly94 shown is not observed in the other three copies of the interface. (B) The potential hydrogen bonds made by Arg135 side chain atoms are not observed in two of the four copies of the interface; there is a potential hydrogen bond between Arg139 and chymotrypsin Ser92 instead. The position of the sulfate ion (Sul) is shown and its hydrogen bonding interactions with surrounding residues are indicated. The side chains of ecotin Arg139 and chymotrypsin Tyr94 have been omitted for clarity. Figure prepared using PyMOL (www.pymol.org).

Figure 5. Alignment of amino acid sequences of Yersinia (B1JSA0) and E. coli (P23827) ecotin. Alignment was carried out using the Clustal W sequence alignment program (34).
## Table 1. Crystallographic data for the ecotin-chymotrypsin complex

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| **Cell dimensions** | a=97.31 Å, b= 48.28 Å, c=174.64 Å  
|                  | α=90˚  
| **Space group (monoclinic) (monoclinic)** | P2₁  
| **Resolution range (Å)** | 50.0 - 2.74  
| **Completeness (%)** | 97.5 (96.9)  
| **Number of reflections** | 273,325  
| **Number of unique reflections** | 42,247  
| **Redundancy (%)** | 3.2 (3.2)  
| **I/σI** | 13.3 (5.5)  
| **R<sub>merge</sub> (%)<sup>a</sup>**<sub>merge</sub> (%)<sup>a</sup> | 7.6 (21.3)  
| **R<sub>cryst</sub> (%)<sup>b</sup>**<sub>cryst</sub> (%)<sup>b</sup> | 24.4  
| **R<sub>free</sub>(%)<sup>c</sup>**<sub>free</sub>(%)<sup>c</sup> | 31.4  
| **Wilson B-factor (Å²)-factor (Å²)** | 54.6  
| **Average B-factors (Å²)** |          |
| Chymotrypsin C1 main chain/side chain | 30.2/30.0  
| Chymotrypsin C2 main chain/side chain | 34.0/34.1  
| Chymotrypsin C3 main chain/side chain | 34.6/34.8  
| Chymotrypsin C4 main chain/side chain | 33.2/33.2  
| Ecotin E1 main chain/side chain | 38.8/39.3  
| Ecotin E2 main chain/side chain | 42.5/42.5  
| Ecotin E3 main chain/side chain | 35.8/36.2  
| Ecotin E4 main chain/side chain<sup>2)</sup> | 44.0/44.5  
| **rmsd from ideal values** |          |
| bonds (Å) | 0.008  
| angles (deg.) | 1.084  

Numbers in parentheses are for the upper resolution shell (2.84 - 2.74 Å), where appropriate.

<sup>a</sup><sub>R</sub><sub>merge</sub> = Σ<sub>hkl</sub>Σ<sub>i</sub>{|I(hkl)−⟨I(hkl)⟩|/Σ<sub>hkl</sub>Σ<sub>i</sub>I(hkl)}, where ⟨I⟩ is the averaged intensity of the i observations of reflection hkl.

<sup>b</sup><sub>R</sub><sub>cryst</sub> = Σ|F<sub>o</sub>|−|F<sub>c</sub>| /Σ|F<sub>o</sub>|, where F<sub>o</sub> and F<sub>c</sub> are observed and calculated structure factors, respectively.

<sup>c</sup><sub>R</sub><sub>free</sub> is equal to <sub>R<sub>cryst</sub></sub> for a random set of reflections (5.1%) not used in refinement (18).
Table 2. Details of interfaces within the asymmetric unit

<table>
<thead>
<tr>
<th>Interface</th>
<th>Total buried surface area (Å²)</th>
<th>Solvation free energy gain upon complex formation (kcal/mol)</th>
<th>Number of hydrogen bonds</th>
<th>Number of van der Waals contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1/E2 dimer</td>
<td>3142</td>
<td>-12.6</td>
<td>20</td>
<td>174</td>
</tr>
<tr>
<td>E3/E4 dimer</td>
<td>2982</td>
<td>-12.1</td>
<td>18</td>
<td>161</td>
</tr>
<tr>
<td>Average</td>
<td>3062</td>
<td>-12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1/C1</td>
<td>1924</td>
<td>-14.3</td>
<td>11</td>
<td>102</td>
</tr>
<tr>
<td>E2/C2</td>
<td>1945</td>
<td>-13.0</td>
<td>9</td>
<td>126</td>
</tr>
<tr>
<td>E3/C3</td>
<td>1923</td>
<td>-16.7</td>
<td>11</td>
<td>109</td>
</tr>
<tr>
<td>E4/C4</td>
<td>1917</td>
<td>-16.2</td>
<td>10</td>
<td>112</td>
</tr>
<tr>
<td>Average</td>
<td>1927</td>
<td>-15.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1/C2</td>
<td>1051</td>
<td>-3.3</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>E2/C1</td>
<td>1064</td>
<td>-2.2</td>
<td>4</td>
<td>56</td>
</tr>
<tr>
<td>E3/C4</td>
<td>1109</td>
<td>-0.7</td>
<td>7</td>
<td>69</td>
</tr>
<tr>
<td>E4/C3</td>
<td>1077</td>
<td>-0.9</td>
<td>6</td>
<td>65</td>
</tr>
<tr>
<td>Average</td>
<td>1076</td>
<td>-1.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Buried surface areas and solvation free energy gains were calculated using PISA (21). Hydrogen bond interactions were identified with the program HBPLUS (33). Contact distances were calculated using the program CONTACT (17) and the maximum allowed contact distances for van der Waals contacts are C-C, 4.1 Å; C-N, 3.8 Å; C-O, 3.7 Å; O-O, 3.3 Å; O-N, 3.4 Å; N-N, 3.4 Å; C-S, 4.1 Å; O-S, 3.7 Å; N-S, 3.8 Å.
Figure 4
Figure 5

| P23827 | QPLEKIAFYQPQAEGKMKROQLTPQEDSTLKVELLIGQTLVEDCNLHRLGGKLENKTL | 84 |
|        | QPLEKIAFYQPQAEGKMROQLTPQEDSTLKVELLIGQTLVEDCNLHRLGGKLENKTL |
| B1JSA0 | QPLEKIAFYQPQAEGMSQVFIFLEPQKDESFKVELLIGKTLNVDCHNMLGGKLENKTL | 90 |

| P23827 | EGWGYDYYVFDKVSSPVSTMMACP-DGKKEKKFVTAYLGDAGMLRYNSKLPIVVYTPDNV | 143 |
|        | EGWGYDYYVFDKVSSPVSTMMACP-DGKKEKKFVTAYLGDAGMLRYNSKLPIVVYTPDNV |
| B1JSA0 | SGWGFYLFYMDKISQPASTMMACPEDSKPOKVFKVTANLGDAMQRYNSKLPIVVYTPDNV | 150 |

| P23827 | DVKYRIVWAKXIDNATVR | 162 |
|        | DVKYRIVWAKXIDNATVR |
| B1JSA0 | EVKYRIWEAGEDIRSAQVK | 169 |

+VKYR+W+A E I +A V+