Structure of a bacterial toxin-activating acyltransferase

Nicholas P Greene, Allister Crow*, Colin Hughes*[†] *and Vassilis Koronakis*

Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, United Kingdom.

*These authors contributed equally to this work.

[†]Author to whom correspondence should be addressed.

CLASSIFICATION: Biological Sciences, Microbiology.

KEYWORDS

Hemolysin, acyltransferase, post-translational modification, crystallography, acyl carrier protein.

ABBREVIATIONS

TAAT, toxin-activating acyltransferase.

ACP, acyl carrier protein.

SAD, single-wavelength anomalous dispersion.

SAXS, small angle x-ray scattering.

SDS-PAGE, sodium dodecyl sulfate – polyacrylamide gel electrophoresis.

GNAT, Gcn5-like N-acetyl transferase.

NEM, N-ethylmaleimide.

ABSTRACT

Secreted pore-forming toxins of pathogenic Gram-negative bacteria such as *Escherichia coli* hemolysin (HlyA) insert into host cell membranes to subvert signal transduction and induce apoptosis and cell lysis. Unusually, these toxins are synthesized in an inactive form that requires post-translational activation in the bacterial cytosol. We have previously shown that the activation mechanism is an acylation event directed by a specialised acyl-transferase that uses acyl carrier protein (ACP) to covalently link fatty acids, via an amide bond, to specific internal lysine residues of the protoxin. We now reveal the 2.15Å resolution X-ray structure of the 172 amino acid ApxC, a toxin-activating acyl-transferase (TAAT) from pathogenic *Actinobacillus pleuropneumoniae*. This shows that bacterial TAATs are a structurally homologous family which, despite indiscernible sequence similarity, form a distinct branch of the GNAT superfamily of enzymes that typically use acyl-CoA to modify diverse bacterial, archaeal and eukaryotic substrates. A combination of structural analysis, small angle X-ray scattering, mutagenesis and cross-linking defined the solution state of TAATs, with inter-monomer interactions mediated by an N-terminal α-helix. Superposition of ApxC with substrate-bound GNATs, and assay of toxin activation and binding of acyl-ACP and protoxin peptide substrates by mutated ApxC variants indicates the enzyme active site to be a deep surface groove.

SIGNIFICANCE

Secreted pore-forming toxins of pathogenic bacteria such as *E.coli* and *Bordetella pertussis* insert into cell membranes to subvert signaling and cause cell death, facilitating infection of human and animal hosts. These toxins require a unique activation step before secretion, the covalent linkage of lipid groups to specific lysines of the inactive protoxin, directed by a specialised toxin-activating acyl transferase (TAAT). Here, we present the TAAT crystal structure, the soluble dimeric topology and likely active site, revealing that despite no discernible sequence similarity TAATs are a structurally and functionally distinct group of the GNAT superfamily of modifying enzymes. Our findings open the way to further understanding of the unique toxin activation, and the possibility of inhibiting toxin action.

\body

INTRODUCTION

Pathogenic bacteria secrete pore-forming protein toxins (PFTs) that target tissue and immune cell membranes to aid colonization and survival during infections, subvert cell signaling, induce apoptosis and promote cell lysis (1–8). Among Gram negative bacteria, large PFTs are secreted by pathogenic species of *Pasteurella*, *Actinobacillus*, *Proteus, Morganella, Moraxella* and *Bordetella*, exemplified by the 110kDa hemolysin (HlyA) of uropathogenic and enterohemorrhagic *E.coli*. These toxins play important roles in cystitis and pyelonephritis, hemorrhagic intestinal disease, periodontitis, pneumonia, septicemia, whooping cough and wound infections (4), and unusually they are made as an inactive protoxin, requiring post-translational activation before export (9–12).

Reconstituting the toxin activation reaction *in vitro* some time ago demonstrated that the essential modification is a novel fatty acid acylation, effected by a specialised coexpressed toxin-activating acyltransferase, in *E.coli* HlyC, that uses acyl-acyl carrier protein (acyl-ACP) as the fatty acid donor (4, 13, 14). The acyltransferase does not share significant sequence identity with other bacterial and eukaryotic enzymes and cellular acyltransferases from either the host or pathogen cannot substitute for HlyC in toxin activation. HlyC independently binds two separate 50-80 amino acid transferase recognition domains (15) each encompassing one of the internal target lysines K564 and K690 of *E.coli* protoxin HlyA, which are acylated by amide linkage, heterogeneously with fatty acids containing 14, 15 and 17 carbon chains (16, 17). Loss of the HlyC binding domain or substitution of protoxin K564 and K690, prevents fatty acyl modification and abrogates all toxin activity (14) as does loss of the transferase (18).

Acylation is essential to the entire family of pore-forming toxins, *B.pertussis* proCyaA lysine acylation has also been demonstrated (19), and the toxin-activating acyltransferases (which we now call TAATs) have high sequence similarity and cross-activate other protoxins (4, 20–22). The TAAT activation mechanism is seemingly unique, and extensive site-directed mutagenesis has so far only identified a single potentially catalytic residue, His23 of HlyC (23–25). Structural information is essential to understand the toxin activation mechanism, and assess TAATs as a potential target for developing anti-virulence compounds that do not affect the host commensal flora. Here we determine the TAAT crystal structure, solution state, and likely active site.

RESULTS

The ApxC structure reveals TAATs as a distinct, closely conserved branch of the GNAT superfamily.

Multiple TAAT genes were cloned and expressed in *E.coli* and where possible TAAT proteins were purified and screened for their amenability to crystallisation. This yielded the crystal structure of the 172 amino acid *Actinobacillus pleuropneumoniae* ApxIC (hereafter ApxC) at 2.15Å resolution using the selenomethionine SAD method of phase determination. Crystallographic data and refinement

statistics are given in **Table S1** and the structure is shown in **Figure 1**. Crystals belong to space group $P2_12_12_1$ with 4 molecules per asymmetric unit (**Fig. 1a**). The four monomers are very similar and can be superposed pairwise with a mean rmsd of 1.7Å (Cα positions, residues 4-170). Structural diversity is largely limited to the last 7 residues of the C-terminus (mean rmsd for residues 5-165 is 0.84Å), and is not expected to be of physiological relevance as this region is not well-conserved among TAATs. Each ApxC monomer is composed of a central five-membered β-sheet that is decorated by six α-helices (**Fig. 1b**). A distinctive split in the sheet is located between strands β3 and β4 which are the only strands that run parallel with one another. This split β-sheet is a key feature of the underlying fold, establishing a deep groove in the enzyme surface.

The overall fold of ApxC is clearly related to the Gcn5-like N-Acetyl Transferase (GNAT) superfamily, confirmed using structural-similarity searches with Dali (26). GNAT superfamily members mediate a diverse range of prokaryotic and eukaryotic processes ranging over antibiotic detoxification, circadian rhythm regulation and polyketide biosynthesis (27, 28). The closest structural match is the GNAT-like domain of CurA (rmsd 2.8Å, Dali Z score=8.6) (29), but robust structural similarity is also evident with other GNATs including the histone acetyl transferases (HATs), Gcn5 (rmsd 3.2Å, Z=7.8) (35), p300/CBP (rmsd 3.1Å, Z=7.9) (30), Myst (rmsd 4.0Å, Z=3.1) (31), and acyl-amino acid synthase FeeM (rmsd 3.6Å, Z=6.6) (32). The ApxC structure is nonetheless distinguished from all other GNATs by the absence of a helical element that typically interacts with the planar nucleotide portion of CoA - this would be located between ApxC residues 122 and 129 (**Fig. 1b,** *teal*), and the CoA diphosphate binding motif ("motif A" in GNATs (27)), that would otherwise be located between strand β3 and helix α3 (residues 97-99, **Fig. 1b,** *orange*). ApxC appears to be the only GNAT that lacks both of these two features, indeed the helix that is missing in ApxC is so uniformly conserved among GNATs that, alongside the split β-sheet, it is considered a defining feature of the GNAT fold (27, 28). In addition to the absence of CoA-interacting elements, the long (25 residue) α-helix at the extreme C-terminus of ApxC (**Fig. 1b,** *violet*) is absent from all known GNAT structures, and a short helix and loop (residues 74 – 86, **Fig. 1b,** *green*) is present only in CurA (29). The close structural relationship between ApxC and members of the GNAT superfamily is remarkable given that there is no primary sequence similarity between these proteins. Even structure-driven sequence alignments produced with Dali reveal just 6.4% sequence identity between ApxC and its closest structural match, the CurA GNAT domain, barely above that expected by chance (5%).

In contrast to the low sequence identity between TAATs and other GNATs, the TAAT family itself is well conserved, illustrated by the multiple sequence alignment in **Figure 2**. ApxC shares 70% primary sequence identity with the best studied TAAT, *E.coli* HlyC, and can directly substitute for HlyC in activation of the *E.coli* hemolysin toxin *in vivo* (**Figure S1**), suggesting the ApxC structure is representative of the TAAT family and allowing us to construct high confidence homology models of homologues *E.coli* HlyC and *Bordetella pertussis* CyaC (**coordinates available in SI**).

TAATs adopt a dimeric arrangement in solution, mediated by an N-terminal helical interface. Gel filtration experiments using recombinant ApxC indicate an apparent molecular weight (42kDa) that is twice that expected for the monomer (21kDa) (**Figure S2**). Examination of the crystal packing in the ApxC structure shows four arrangements that could conceivably represent this dimer (Dimers I to IV in **Figure 3a**). Dimer I, underpinned by the unique C-terminal helix and Dimer III which appears twice in the asymmetric unit (**Figure S3**) were attractive possibilities but it was not obvious by inspection which interface reflects the physiological dimer and which are crystal contacts.

We assessed the extent and quality of each of the four dimer interfaces by measuring the buried surface area and calculating the predicted stability using PISA (33) (**Fig. 3a**). Dimer IV had the lowest buried surface area and was predicted to be unstable but Dimers I, III and IV had substantial buried surface area and were all predicted to be stable. Analysis of scattering data from small angle X-ray scattering (SAXS) (**Fig. 3b**) indicated that the particle has an estimated molecular weight of ~38.8kDa, a 28Å radius of gyration and a probable maximum dimension of ~77Å. *Ab initio* shape reconstructions of the solution particle were consistent with a 2-fold symmetric bilobal globular dimer (**Fig. 3b**). Comparison of the scattering data with theoretical profiles predicted for each dimer was used to assess which arrangement most closely resembles the solution particle (**Fig. 3c**). Both Dimer I and Dimer III have low χ^2 values indicating a good fit between prediction and experiment while Dimers II and IV do not. Inspection of the predicted P(r) profiles (**Fig. 3c**, black lines) shows that both Dimers I and III are close to the experimentally-derived profile (blue lines) consistent with bilobal structures with a maximum dimension of ~80 Å, while Dimers II and especially IV deviate significantly suggesting that the maximum dimensions of these dimers are respectively 10Å and 15Å too short to be consistent with the data. Taken together, the surface areas, predicted stabilities and SAXS data excluded Dimers II and IV but did not satisfactorily discern between Dimers I and III.

We therefore applied two further molecular genetic experimental approaches. Firstly, we conducted *in vitro* cross-linking experiments with the aim of covalently trapping the physiological dimer using experimentally-engineered cysteine residues placed independently at each potential dimerization interface. Secondly, we used a bacterial 2-hybrid system (34) to monitor the inter-monomer interaction *in vivo* after introducing disruptive mutations to each putative interface. The interface structures, genetic interventions and results are presented in **Figure 4**. Both Dimers I and III are 2-fold rotationally symmetric, so it was possible to introduce single cysteines close to the relevant rotation axis to generate potential cross-link sites, namely Gly12Cys within the N-terminal helix to specifically trap Dimer III, and Lys160Cys within the C-terminal helix to trap Dimer I (**Fig. 4a**). Cross-linking of the dimer was successful with Gly12Cys, but not Lys160Cys,

and no cross-linked dimer was generated by the wild type protein, nor when solvent accessible cysteines were pre-blocked with NEM (**Fig. 4b**). In parallel, we used a bacterial 2-hybrid assay (34) in which wildtype ApxC dimerization results in reporter activity (β-galactosidase activity, blue spot, top box **Fig. 4d**). Mutations were engineered to disrupt either the N-terminal helix within the Dimer III interface (Gly12Ala, Gly12Glu, Ala15Glu, and Ala19Glu substitutions, **Fig. 4c**) or the intertwined C-termini of Dimer I by a 10 amino acid deletion (**Fig. 4e**), all variant fusion proteins expressed comparably (**Figure S4**). The Gly12Glu and Ala15Glu mutation disrupted the dimer (white spots, **Fig. 4d**), while deletion of the C-terminus had no effect (**Fig. 4f**). Taken together these experiments unambiguously define Dimer III as the state of the enzyme in solution.

Identification of the TAAT active site within a deep surface groove.

The structural similarity between TAATs and the GNATs provoked a search for shared catalytic residues in structural superimpositions of ApxC with key members of the Gcn5 (35), Hat1 (30), Myst (36), p300/CBP (30), FeeM(32) and CurA (29) GNAT subfamilies. Statistics for the structural superpositions are given in **Table S2** alongside summaries of the expected active site residues and the ApxC residues to which they most closely align. Our search confirmed that TAATs do not contain catalytic residues established for other GNATs including the buried glutamates found in the Gcn5 and Hat1 families (27, 28), the cysteine and glutamate found in Myst (36), or the phenylalanine and tyrosine residues thought responsible for substrate alignment in the CBP/p300 subfamily (30). To assess potential substrate-binding sites, ApxC was also compared with substrate-bound structures of the GNATs Gcn5 (35) and FeeM (32) (**Figure 5**). Gcn5, the archetypal member of GNAT superfamily is a histone acetyltransferase that catalyses the acetyl CoA dependent acylation of a sequence-internal histone lysine residue. The crystal structure of the ternary complex (35) formed between Gcn5, CoA and a peptide representing the histone tail superposes well with ApxC with parts of both substrates overlaying portions of the TAAT active site groove (**Fig. 5a**). FeeM catalyses the acylation of tyrosine using FeeL, an ACP-like protein, as acyl donor. Structural superposition of FeeM with ApxC places the acyl chain of the N-lauroyl tyrosine product (co-crystallised with FeeM) within a hydrophobic portion of the TAAT surface groove (**Fig. 5b**). These superpositions suggest that the deep groove observed for ApxC is the likely substrate binding site.

To confirm the location of the active site and to identify key residues, we generated alanine mutants of residues lining the groove, in particular targeting polar residues conserved in the TAAT family (**Fig. 2**). These included His24 for which the equivalent residue in *E.coli* TAAT HlyC has been shown to be important (24), Ser21 and Asp93 which together with His24 have a three dimensional arrangement reminiscent of catalytic triads found in non-GNAT acyltransferases and proteases). The structural context of these targeted residues is shown in **Fig. 6a,** and relative TAAT activities of ApxC alanine-variants assessed by erythrocyte lysis of variant-activated toxin

are indicated in **Fig. 6b**. The Ser21Ala substitution caused only a marginal (21%) reduction in TAAT activity, inconsistent with a role as a nucleophile within a catalytic triad. A role for the TAAT family conserved cysteine residue can similarly be excluded as the ApxC structure reveals Cys58 is completely buried within the core of the protein where it is inaccessible to substrates (**Fig S5a, b**). In agreement with this, the Cys58Ala variant has wild type levels of activity (**Fig. S5c**). In contrast, both Asp93 and Arg121 alanine variants were profoundly affected by mutation with TAAT activities less than 1% of the wild type. The His24 and Asn35 variants were also significantly impaired, although not to the same degree, with measurable activities of c.5% of wild type.

To probe the roles of these four important residues, we performed bacterial 2-hybrid experiments to assess their ability to engage with TAAT substrates (i.e. the protoxin and ACP). All four alanine-substituted variants were able to dimerise efficiently indicating that each is properly folded (**Fig. 6c**, *left*), and all engaged with a 58 amino acid peptide (residues 679-736 of the protoxin) known to be the recognition sequence for TAAT spanning K690 (15) (**Fig. 6c**, *centre*). Asn35, Asp93 and His24 alanine variants were also able to bind ACP, but the Arg121Ala mutant was not (**Fig. 6c**, *right*) despite being expressed at comparable level to the wild type (**Fig. 6b**). We therefore anticipate Arg121 mediates interaction with ACP while the remaining essential residues, His24, Asn35, and Asp93, are likely to have roles in catalysis. The putative active sites encompassing these residues in the surface groove are on opposing faces within the dimer, not in close juxtaposition (**Fig. 6d**).

DISCUSSION

We have described the high resolution crystal structure (**Fig. 1**), solution structure (**Fig. 3b**), dimerisation interface (**Figs. 3 & 4**) and active site location (**Fig. 6**) of ApxC, revealing the close-knit family of toxin-activating acyltransferases (TAATs) as a new subdivision of the GNAT superfamily. While TAATs retain the characteristic GNAT split in the central β-sheet that underlies the putative substrate-binding groove, they lack the structural motifs typically associated with CoA binding, including both the nucleotide-interacting helix and diphosphate-binding motif that are considered core elements of the characteristic GNAT fold. A long C-terminal α-helix and small loop-helix feature further distinguish TAATs from other GNATs (**Fig. 1**).

In vitro SAXS (**Fig. 3**) and *in vivo* two-hybrid analyses (**Fig. 4**) confirmed indications from gel filtration (previously with HlyC (18) and here with ApxC) that TAATs are dimers in solution. Analysis of biophysically tenable dimers abstracted from the crystal structure using a combination of experimental solution scattering data, targeted mutagenic disruption and cross-link locking of potential inter-monomer contacts, firmly established the dimer topology (**Fig. 3a**, Dimer III). The dimer interface is largely hydrophobic, with the close approach of the N-terminal helices facilitated by the small side chains of Gly12 and Ala15. Consequently, the proposed active sites of the monomers, each located within a surface groove, are topologically distinct within the context of the

dimer and unlikely to act in concert (**Fig. 6d**).

The lack of typical GNAT CoA-binding features in TAATs is understandable from a functional perspective as TAATs use ACP as the substrate acyl donor and not CoA. However, FeeM is also dependent on an ACP-like acyl donor (FeeL) yet retains the GNAT-typical CoA-interacting helix. The interaction between FeeM and FeeL has been partially mapped by mutation and involves the helical region that is absent in TAATs (32) suggesting the TAAT:ACP interaction must be distinct from that of FeeM:FeeL. The only other structurally characterised GNAT known to interact with ACP is CurA, a large multidomain protein involved in polyketide synthesis. The CurA GNAT domain interacts with both CoA and an ACP-like internal domain facilitating acyl transfer between these substrates (29). CoA binding in CurA occurs in typical GNAT fashion while a distinct ACP binding site has been proposed to lie in a separate tunnel that connects directly to the CoA binding site. A similar binding mode is possible in TAATs as the proposed ACP-binding tunnel of CurA partially overlaps the ApxC surface groove.

Structure-led mutagenesis of the TAAT surface groove, the putative active site, revealed two residues, Asp93 and Arg121, as essential to function and Asn35 and His24 as important. Mutation to alanine did not impair dimerisation, and while all four alanine variants still interacted with the target protoxin, mutation of Arg121 uniquely abrogated binding of ACP. In other unrelated ACP-binding enzymes positively charged residues mediate direct interactions with phosphates of the ACP phosphopantetheine group (37) or the negatively charged residues of ACP (38), and so it seems likely that Arg121 plays a similar role in the TAAT:ACP interaction. For Asn35, the chemistry of the asparagine side chain precludes a role in acid-base catalysis and amide group is unlikely to be a nucleophilic, but it may have a role in orientating substrates or other essential residues e.g. Asp93 to which it is hydrogen bonded.

His24 and Asp93 are both candidates to be true catalytic residues and plausible roles within the context of the two most likely acyl transfer schemes are outlined in **Figure S6**. In the first scheme, a ternary complex is formed between the enzyme and both substrates facilitating direct nucleophilic attack of the protoxin target lysine on the acyl-ACP thioester with subsequent expulsion of ACP (a “direct attack” mechanism, **Fig. S6a**). In the second, the reaction proceeds in two steps through formation of a covalent acyl-enzyme intermediate that is resolved by attack of the toxin receptor lysine (“covalent catalysis”, **Fig. S6b**). In either mechanistic scheme, analogy to other GNATs, particularly histone acetyltransferases such as Gcn5 and Hat1, suggests Asp93 acts as general base to deprotonate the receptor lysine of the protoxin, and in the specific case of covalent catalysis it could additionally deprotonate the catalytic nucleophile. His24 would then serve as general acid protonating the ACP thiolate leaving group in the direct attack scheme, or else as the enzymatic nucleophile in the covalent catalysis scheme.

That TAATs represent a novel subfamily of GNATs is an important finding as kinetically characterised GNATs appear to use a ternary complex to facilitate transfer of acyl groups from one

substrate to the other giving strong support for a ternary complex (direct attack) based mechanism (27, 35, 39, 40). This agrees with early kinetic experiments on HlyC (41), but is at variance with other results favouring sequential binary interactions between enzyme and substrate (42), and observations of acylenzyme formation inferred from experiments with radiolabelled fatty acids (23, 24). Incubation of TAAT HlyC with acyl-ACP does yield acylated enzyme and is dependent on integrity of the active site histidine (equivalent to ApxC His24) (23). However, direct histidine modification has not been demonstrated and in other GNATs internal enzyme acylation is suggested to be a self-regulatory feature (36, 43), so it is not clear whether the acylated TAAT is an acyl transfer intermediate, regulatory state, or artefact of acyl donor interaction in the absence of the acyl acceptor.

In support of ternary complex formation, the TAAT surface groove is of comparable size to that of Gcn5 and so probably big enough to accommodate both the acylated phosphopantetheine arm of acyl-ACP and the protoxin lysine, if not also a substantial part of unfolded protoxin. Based on the superposition of TAATs with other GNATs (**Fig. 5**) and our identification of the TAAT catalytic residues (**Fig. 6**) we illustrate how TAATs could bring together the acyl- ACP and toxin substrates to effect toxin activation in **Fig. 7**. The central location of the active site residues within the groove (*yellow*, **Fig. 7b)** hints that each substrate could be bound on opposing faces of the enzyme with acyl transfer chemistry occurring at the crux of the split β-sheet. In support of this, the ACP-interacting arginine (Arg121 in ApxC) is located on the same face of the enzyme as the portion of the active site groove that accommodates a lauryl acyl chain in FeeM (**Fig. 7a**) suggesting the likely ACP binding site (**Fig. 7b,** *green*) and this is the opposite face to which the histone tail locates in the Gcn5 ternary complex (**Fig. 7a**, *blue*) suggesting a possible protoxin binding site, (**Fig. 7b**, *blue*). If TAATs do use a ternary complex it seems likely they would also use the direct attack mechanism typical of other GNATs. Otherwise, TAATs will be of significant interest as a truly unique subset of the GNAT superfamily.

Elucidation of the TAAT protein structure, solution oligomer and active site represents a significant step forwards towards understanding how bacterial pore-forming toxins are activated. The structural and possibly functional distinction between TAATs and all known GNAT families also identifies these enzymes as possible targets for developing new anti-virulence compounds.

METHODS

Cloning and site-directed mutagenesis.

The *apxIC* gene was PCR-amplified from *Actinobacillus pleuropneumoniae* (ATCC27088) genomic DNA using primers ApxC QE80 F and ApxC QE80 R (all primer sequences are given in **Table S3**) engineered to encode restriction sites in the 5'-ends. The PCR product was digested with BamHI and HindIII and ligated into pQE80 (Qiagen). Site-specific mutations were introduced using QuikChange (Stratagene). Clones from PCR-amplified DNA were sequenced (Source Bioscience).

TAAT genes from uropathogenic *E.coli,* enterohemorrhagic *E.coli*, *Bordetella pertussis, Mannheimia hemolytica*, *Aggregatibacter actinomycetemcomitans,* and *Aeromonas hydrophila* were cloned with a similar strategy. For bacterial 2-hybrid ApxC or variant was amplified with primers ApxC HindIII F/ApxC SacI REV (ApxC HindIII F/ApxC 161 R for the ∆162-172 construct), digested with HindIII and SacI and cloned into pUT18 and pKNT25 resulting in the adenylate cyclase fragment being placed at the C-terminus of ApxC. Similarly *E.coli* ACP and a fragment corresponding to HlyA residues 679-736 were amplified with primer pairs ACP F/ACP R and HlyA 679 F/HlyA 736 R respectively, digested with HindIII-SacI and cloned into pUT18 digested with the same enzymes.

ApxC protein expression and purification.

N-terminally His-tagged ApxC was produced in *E. coli* C41 cells (44) bearing plasmid pQE80/ApxC, grown in 2TY medium at 30 °C until A_{600} 0.6 then the temperature was reduced to 18 °C and expression induced with 0.1mM IPTG for 16h. Cells were harvested by centrifugation (10,000 *g*) and resuspended in 50mM HEPES pH 7.5, 400mM NaCl, 5% (v/v) glycerol supplemented with Complete EDTA-free protease inhibitor mixture (Roche) and broken by two passages through a cell disruptor (30,000 psi). Unbroken cells were removed by ultracentrifugation at 150,000 *g* for 1h at 4°C. The supernatant was supplemented with 4mM Imidazole and added to Profinity IMAC resin (Biorad) for 1h at 4°C. Protein-bound resin was washed with 25mM HEPES pH 7.5, 500mM NaCl, 8mM Imidazole, 0.1% Triton X-100, then with the same buffer without Triton X-100, and eluted with 25mM HEPES pH 7.5, 200mM NaCl, 250mM Imidazole. Protein fractions were pooled, concentrated and loaded onto a Superdex S75 column equilibrated in 25mM HEPES pH 7.5, 200mM NaCl. Peak fractions were concentrated to 5mg/ml in an Amicon 10kDa centrifugal filter device (Millipore). For phase determination, SeMet-labelled ApxC was produced in *E.coli* B834 transformed with pQE80/ApxC. Cells grown in M9 minimal medium supplemented with 50 mg/ml methionine to A_{600} 0.8, were resuspended in minimal media lacking methionine for 20min and selenomethionine added to 50mg/ml final concentration. After another 20min, expression was induced by IPTG as before, and SeMet-labelled protein purified as unlabelled protein except that 1mM TCEP (Merck) was added to buffers. Protein identity and SeMet incorporation were verified by mass spectrometry (PNAC, Cambridge Department of Biochemistry).

Crystallisation and structure determination.

ApxC crystals were obtained via the sitting-drop vapour diffusion method using crystallisation reagent composed of 32% (w/v) PEG 300, 100mM Phosphate/Citrate Buffer pH 4.4, and a protein solution of 5mg/ml ApxC in 150mM NaCl, 20mM HEPES pH 7.5. Sitting drops were formed by mixing 2µL protein solution with 1µL crystallization reagent and equilibration over a 500µL volume of the crystallization reagent at 15°C in 24-well plates. Crystals appeared after 3 days and were

harvested a few days later. For freezing, crystals were transferred into a cryoprotectant solution composed of 20mM Phosphate/Citrate buffer, 50mM NaCl, 7.5% (w/v) PEG 300 and 10% (v/v) glycerol, then to the same solution containing 20 % (v/v) glycerol before harvesting with a cryoloop and flash-freezing in liquid nitrogen. Crystals of the selenomethionine derivative were prepared similarly, although were typically much smaller than those of the native protein. X-ray diffraction data were collected on beamlines I02 and I04 of Diamond Light Source, UK. Structure determination utilised the Se-SAD method of phase determination aided by programs from both the PHENIX (45) and CCP4 software suites (46). X-ray data were indexed and integrated using MOSFLM (47) and scaled using SCALA. A partial substructure of Selenium sites was identified using HYSS (48) and experimental phases calculated with PHASER (49). Density modification using PARROT (50) benefitted from identification of a 2-fold non-crystallographic symmetry relating a subset of the heavy atom sites. Manual inspection of the density-modified maps identified further selenium sites and derivation of the likely 4-fold NCS operators relating each of the four monomers in the asymmetric unit. After further density modification using 4-fold NCS averaging, an initial selenoApxC model was built with BUCCANEER (51) and completed with iterative rounds of manual model-building with COOT (52) and refinement with REFMAC (53) using NCS restraints. Refinement was switched to an isomorphous, high-resolution, native dataset maintaining the same 'free' reflection set, with NCS-restraints on the monomers relaxed and solvent components such as ordered water molecules and crystallization reagents added to complete the crystallographic model. Further model validation used PROCHECK (54), and RAMPAGE (55). The final model comprises 4 molecules of ApxC, 219 waters and a citrate molecule. Atomic coordinates and structure factors are deposited in the Protein Data Bank, accession code 4WHN.

Structure analysis, sequence alignments and homology modelling.

Buried surface areas were calculated using AREAIMOL (56) with a solvent probe radius of 1.2Å. Predicted dimer stability was assessed with PISA (33). Root mean square deviations (rmsds) were calculated over all C_α atoms after alignment using superpose in the CCP4 suite (57). Figures were prepared in Pymol (58). For homology modelling, a single monomer of ApxC was used as a template on which to impose the sequence of other TAATs using the Phyre2 server in sequence threading mode (59). For multiple sequence alignments, amino acid sequences were processed using Clustal Omega (60) and conserved residues mapped with ESPript (61). Further editing to add the secondary structure elements and essential residues was performed manually using the GNU Image Manipulation Programme (GIMP).

Small angle X-ray scattering (SAXS) analysis.

Small angle X-ray scattering was conducted at the PETRA P12 SAXS beamline in Hamburg, Germany (Deutsches Elektronen-Synchrotron – DESY) using purified ApxC protein in buffer 20mM

HEPES pH 7.5, 150mM NaCl. Data used in final analysis is the average of 4 separate samples (each the average of 10-20 frames) collected using 56μM ApxC. Computational analysis of the data used the ATSAS package of programs (62). Inspection and characterisation of the data used PRIMUS (63) and GNOM (64), while *ab initio* shape reconstruction used DAMMIN (65). The final particle was constructed by averaging 20 independent models with DAMAVER (66) before filtering with DAMFILT (66). The final model also used enforced P2 symmetry, but runs without symmetry restraints produced similar results. CRYSOL (67) was used to compare predicted scattering behaviour of dimers/monomers in our crystal structure with scattering data obtained in solution.

Hemolysis assay of toxin activation.

E. coli MC1061 (TolC positive) cells bearing plasmids expressing the export system HlyBD (LG575 (68)) and HlyA (LG583 (68)) were transformed with pQE80 vector alone or expressing wildtype or mutant ApxC. Bacteria were grown overnight on blood agar plates supplemented with 20μg/mL chloramphenicol and 50μg/mL carbenicillin. LB cultures supplemented with antibiotics to maintain plasmid selection were grown to A_{600} 0.8 and cells removed. Dilutions of supernatant containing secreted HlyA were added to 2% washed horse erythrocytes, 20mM $CaCl_2$, 150mM NaCl and incubated 30min at 42°C. Hemolytic activity was assayed as hemoglobin release at A_{543} after subtracting the absorbance of the vector control. Assays were performed in triplicate, with each absorbance measurement itself the mean of triplicates.

Bacterial two-hybrid assays.

E.coli BTH101 was co-transformed with recombinant pUT18 and pKNT25 plasmids expressing the protein of interest fused to either the T18 or T25 domain of *B.pertussis* CyaA and grown at 30°C. Overnight culture (LB with 100 μg/mL carbenicillin and 50μg/mL kanamycin) grown at 30°C was spotted (2μl) on LB agar containing 0.5mM IPTG and 40μg/mL X-GAL. After overnight incubation at 30°C plates were photographed under white light. β-galactosidase assays were performed as described (69).

Cross-linking.

Cysteine variants of ApxC were purified as for the wildtype protein except that 2mM TCEP was added to all buffers. For cross-linking reactions, TCEP was removed by desalting column and 100μM $CuCl_2$ added to 50μg of protein in 25mM HEPES pH7.5, 150mM NaCl and incubated at 25ºC for 30 minutes before addition of 10mM EDTA (to chelate copper) and 5mM NEM (to block free cysteines). Reactions were analysed by SDS PAGE. For the control reaction (pre-block) samples were pretreated with EDTA and NEM prior to addition of $CuCl_2$.

Western blot analysis.

Samples were resolved on SDS PAGE, transferred to nitrocellulose membrane and immunoblotted with anti-His (Qiagen) or anti-cyaA (Santa Cruz) antisera as appropriate. These were detected with a chemiluminescence kit (GE Healthcare ECL) or fluorescence imaging (Odyssey Licor) respectively.

ACKNOWLEDGEMENTS

We thank staff at Diamond Light Source (UK) and EMBL Hamburg for beam line provision, Dr. Len Packman for mass spec services, Dr Phil Hinchliffe for assisting data collection, and Gouzel Karimova (Institut Pasteur, Paris) for the generous gift of plasmids pKNT25 and pUT18. This work was supported by grants to CH and VK from the UK Medical Research Council and the Wellcome Trust.

AUTHOR CONTRIBUTIONS

NG, AC, CH and VK designed experiments; NG performed molecular biology, microbiological assays, cross-linking and crystallised ApxC; AC solved the structure and performed SAXS; All authors contributed to analysis and writing of the paper; CH and VK supervised the work.

REFERENCES

1. Ulett GC, *et al*. (2013) Uropathogenic *Escherichia coli* virulence and innate immune responses during urinary tract infection. *Curr Opin Microbiol* 16(1):100–7.

2. Gur C, *et al*. (2013) Natural Killer Cell-Mediated Host Defense against Uropathogenic *E. coli* Is Counteracted by Bacterial HemolysinA-Dependent Killing of NK Cells. *Cell Host Microbe* 14(6):664–74.

3. Uhlén P, *et al.* (2000) Alpha-haemolysin of uropathogenic *E. coli* induces Ca^{2+} oscillations in renal epithelial cells. *Nature* 405(6787):694–7.

4. Stanley P, Koronakis V, Hughes C (1998) Acylation of *Escherichia coli* hemolysin: a unique protein lipidation mechanism underlying toxin function. *Microbiol Mol Biol Rev* 62(2):309–33.

5. Wiles TJ, Dhakal BK, Eto DS, Mulvey MA (2008) Inactivation of host Akt/protein kinase B signaling by bacterial pore-forming toxins. *Mol Biol Cell* 19(4):1427–38.

6. Smith YC, Rasmussen SB, Grande KK, Conran RM, O'Brien AD (2008) Hemolysin of uropathogenic *Escherichia coli* evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice. *Infect Immun* 76(7):2978–90.

7. Dhakal BK, Mulvey MA (2012) The UPEC pore-forming toxin α-hemolysin triggers proteolysis of host proteins to disrupt cell adhesion, inflammatory, and survival pathways. *Cell Host Microbe* 11(1):58–69.

8. Nagamatsu K, et al. (2015) Dysregulation of *Escherichia coli* α-hemolysin expression alters the course of acute and persistent urinary tract infection. *Proc Natl Acad Sci U S A* 112(8):E871–880.

9. Koronakis V, Eswaran J, Hughes C (2004) Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annu Rev Biochem* 73:467–89.

10. Koronakis V, Koronakis E, Hughes C (1989) Isolation and analysis of the C-terminal signal directing export of *Escherichia coli* hemolysin protein across both bacterial membranes. *EMBO J* 8(2):595–605.

11. Thanabalu T, Koronakis E, Hughes C, Koronakis V (1998) Substrate-induced assembly of a contiguous channel for protein export from *E.coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J* 17(22):6487–96.

12. Balakrishnan L, Hughes C, Koronakis V (2001) Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by *Escherichia coli*. *J Mol Biol* 313(3):501–10.

13. Issartel JP, Koronakis V, Hughes C (1991) Activation of *Escherichia coli* prohaemolysin to the mature toxin by acyl carrier protein-dependent fatty acylation. *Nature* 351(6329):759–61.

14. Stanley P, Packman LC, Koronakis V, Hughes C (1994) Fatty acylation of two internal lysine residues required for the toxic activity of *Escherichia coli* hemolysin. *Science* 266(5193):1992–6.

15. Stanley P, Koronakis V, Hardie K, Hughes C (1996) Independent interaction of the acyltransferase HlyC with two maturation domains of the *Escherichia coli* toxin HlyA. *Mol Microbiol* 20(4):813–22.

16. Ludwig A, et al. (1996) Analysis of the in vivo activation of hemolysin (HlyA) from *Escherichia coli*. *J Bacteriol* 178(18):5422–30.

17. Lim KB, et al. (2000) *Escherichia coli* alpha-hemolysin (HlyA) is heterogeneously acylated in vivo with 14-, 15-, and 17-carbon fatty acids. *J Biol Chem* 275(47):36698–702.

18. Hardie KR, Issartel JP, Koronakis E, Hughes C, Koronakis V (1991) In vitro activation of *Escherichia coli* prohaemolysin to the mature membrane-targeted toxin requires HlyC and a low molecular-weight cytosolic polypeptide. *Mol Microbiol* 5(7):1669–79.

19. Hackett M, Guo L, Shabanowitz J, Hunt DF, Hewlett EL (1994) Internal lysine palmitoylation in adenylate cyclase toxin from *Bordetella pertussis*. *Science* 266(5184):433–5.

20. Forestier C, Welch RA (1990) Nonreciprocal complementation of the hlyC and lktC genes of the *Escherichia coli* hemolysin and *Pasteurella haemolytica* leukotoxin determinants. *Infect Immun* 58(3):828–832.

21. Gygi D, *et al.* (1990) Isolation of the *Actinobacillus pleuropneumoniae* haemolysin gene and the activation and secretion of the prohaemolysin by the HlyC, HlyB and HlyD proteins of Escherichia coli. *Mol Microbiol* 4(1):123–8.

22. Koronakis V, Cross M, Senior B, Koronakis E, Hughes C (1987) The secreted hemolysins of *Proteus mirabilis*, *Proteus vulgaris*, and *Morganella morganii* are genetically related to each other and to the alpha-hemolysin of *Escherichia coli*. *J Bacteriol* 169(4):1509–15.

23. Trent MS, Worsham LM, Ernst-Fonberg ML (1999) HlyC, the internal protein acyltransferase that activates hemolysin toxin: roles of various conserved residues in enzymatic activity as probed by site-directed mutagenesis. *Biochemistry* 38(29):9541–8.

24. Trent MS, Worsham LM, Ernst-Fonberg ML (1999) HlyC, the internal protein acyltransferase that activates hemolysin toxin: role of conserved histidine, serine, and cysteine residues in enzymatic activity as probed by chemical modification and site-directed mutagenesis. *Biochemistry* 38(11):3433–9.

25. Trent MS, Worsham LM, Ernst-Fonberg ML (1999) HlyC, the internal protein acyltransferase that activates hemolysin toxin: the role of conserved tyrosine and arginine residues in enzymatic activity as probed by chemical modification and site-directed mutagenesis. *Biochemistry* 38(27):8831–8.

26. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545–9.

27. Dyda F, Klein DC, Hickman AB (2000) GCN5-related N-acetyltransferases: a structural overview. *Annu Rev Biophys Biomol Struct* 29:81–103.

28. Vetting MW, *et al*. (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys* 433(1):212–26.

29. Gu L, *et al*. (2007) GNAT-like strategy for polyketide chain initiation. *Science* 318(5852):970–4.

30. Clements A, *et al*. (1999) Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J* 18(13):3521–32.

31. Yan Y, Harper S, Speicher DW, Marmorstein R (2002) The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. *Nat Struct Biol* 9(11):862–9.

32. Van Wagoner RM, Clardy J (2006) FeeM, an N-acyl amino acid synthase from an uncultured soil microbe: structure, mechanism, and acyl carrier protein binding. *Structure* 14(9):1425–35.

33. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–97.

34. Karimova G, Pidoux J, Ullmann A, Ladant D (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* 95(10):5752–6.

35. Rojas JR, *et al*. (1999) Structure of *Tetrahymena* GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* 401(6748):93–8.

36. Sun B, *et al*. (2011) Regulation of the histone acetyltransferase activity of hMOF via autoacetylation of Lys274. *Cell Res* 21(8):1262–6.

37. Masoudi A, Raetz CRH, Zhou P, Pemble IV CW (2013) Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature*.

38. Nguyen C, *et al*. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* 505(7483):427–31.

39. Lau OD, *et al*. (2000) p300/CBP-associated factor histone acetyltransferase processing of a peptide substrate. Kinetic analysis of the catalytic mechanism. *J Biol Chem* 275(29):21953–9.

40. Berndsen CE, Albaugh BN, Tan S, Denu JM (2007) Catalytic mechanism of a MYST family histone acetyltransferase. *Biochemistry* 46(3):623–9.

41. Stanley P, Hyland C, Koronakis V, Hughes C (1999) An ordered reaction mechanism for bacterial toxin acylation by the specialized acyltransferase HlyC: formation of a ternary complex with acylACP and protoxin substrates. *Mol Microbiol* 34(5):887–901.

42. Worsham LM, Trent MS, Earls L, Jolly C, Ernst-Fonberg ML (2001) Insights into the catalytic mechanism of HlyC, the internal protein acyltransferase that activates *Escherichia coli* hemolysin toxin. *Biochemistry* 40(45):13607–16.

43. Yang C, Wu J, Sinha SH, Neveu JM, Zheng YG (2012) Autoacetylation of the MYST lysine acetyltransferase MOF protein. *J Biol Chem* 287(42):34917–26.

44. Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260(3):289–98.

45. Adams PD, *et al*. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–21.

46. Winn MD, *et al*. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–42.

47. Read RJ, Sussman JL eds. (2007) *Evolving Methods for Macromolecular Crystallography* (Springer Netherlands, Dordrecht).

48. Grosse-Kunstleve RW, Adams PD (2003) Substructure search procedures for macromolecular structures. *Acta Crystallogr Sect D* 59(11):1966–1973.

49. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(4):658–674.

50. Cowtan K (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):470–8.

51. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62(Pt 9):1002–11.

52. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D* 66(4):486–501.

53. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–55.

54. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26(2):283–291.

55. Lovell SC, *et al*. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 50(3):437–50.

56. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55(3):379–400.

57. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2256–68.

58. Schrodinger L (2010) *The PyMOL Molecular Graphics System, Version~1.3r1*.

59. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4(3):363–71.

60. Goujon M, *et al*. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38 (Web Server issue):W695–9.

61. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42 (Web Server issue):W320–4.

62. Petoukhov M V., *et al*. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45(2):342–350.

63. Konarev P V., Volkov V V., Sokolova A V., Koch MHJ, Svergun DI (2003) PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36(5):1277–1282.

64. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25(4):495–503.

65. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6):2879–86.

66. Volkov V V., Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36(3):860–864.

67. Svergun D, Barberato C, Koch MHJ (1995) CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr* 28(6):768–773.

68. Mackman N, Nicaud JM, Gray L, Holland IB (1985) Genetical and functional organisation of the *Escherichia coli* haemolysin determinant 2001. *Mol Gen Genet* 201(2):282–8.

69. Battesti A, Bouveret E (2012) The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*. *Methods* 58(4):325–34.

FIGURE LEGENDS

Figure 1: Crystal structure of the toxin-activating acyl transferase ApxC. (**a**) The asymmetric unit contains four monomers of ApxC. An imperfect 2-fold non-crystallographic symmetry axis (dotted red line) relates two pairs of monomers. (**b**) Structure of a single monomer indicating structural features that discern TAATs from other members of the GNAT superfamily.

Figure 2: Multiple sequence alignment for key members of the TAAT family. Secondary structure of ApxC (consensus of the four monomers in the crystal structure) is shown immediately above the alignment of TAAT sequences ApxC (*Actinobacillus pleuropneumoniae*), HlyC *(Uropathogenic Escherichia coli*), AaLktC (*Aggregatibacter actinomycetemcomitans*), MhLktC (*Mannheimia hemolytica*) and CyaC (*Bordetella pertussis*). The α-helices indicated are shown as tubes (*blue*), β-strands by arrows (*green*). Conserved residues are highlighted *red*, and important active site residues identified later (His24, Asn35, Asp93 and Arg121) are shown in *gold*. Features that differentiate TAATs from other GNATs are highlighted using the same colour scheme as in Figure 1b.

Figure 3: X-ray analysis of potential ApxC dimers *in crystallo* and in solution. (**a**) The four dimeric arrangements of ApxC extracted from the crystal structure (I-IV) indicating buried surface area and predicted free energy of dissociation *(*ΔG, positive values indicate favourable association) obtained from PISA, (33). (**b**) Small Angle X-ray Scattering (SAXS) analysis of ApxC in solution. (*i*) The measured SAXS scattering intensity profile is shown in *blue* (56μM ApxC) and the predicted scattering profile of the solution model in *black*, while the solution structure of ApxC, a filtered model derived from the average of 20 models, is shown inset. (*ii*) Distance distribution analysis of each of the candidate ApxC dimers verses the SAXS data. The quality of the fit can be seen by comparing predicted P(r) profile in *black* with the measured profile in *blue*, quantified by the χ^2 value shown inset.

Figure 4: Experimental determination of the ApxC solution dimer. (**a**) Locations of engineered cysteine residues (marked with spheres Gly12, *blue*; Lys160 *orange*) with respect to two dimeric arrangements, (Dimer III, *left*; Dimer I, *right*). (**b**) Cysteine cross-linking of purified wild type (WT) ApxC or cysteine variant. Protein (50μg) was treated with or without 100μM $CuCl_2$ (X-linker) for 30 minutes at 25ºC to catalyse cross-linking of proximal cysteines before blocking free cysteines with 5mM N-ethylmaleimide (NEM). Pre-block reactions were treated with NEM and EDTA prior to addition of cross-linker. (**c**) Locations of Gly12, Ala15 and Ala19 within the N-terminal α-helix at the Dimer III interface. (**d**) Bacterial 2-hybrid assay of homodimerisation of WT ApxC or indicated

variants. Interaction was assessed qualitatively by blue colouration and quantitatively by β-galactosidase activity (right), expressed as a percentage of the wild type ± SD. (**e**) Location of an engineered C-terminal deletion (Δ162-172) in context of Dimer I. (**f**) Bacterial 2-hybrid assay assessing homodimerisation of ApxC (Δ162-172), details as for (d).

Figure 5: Superposition of the TAAT ApxC with Gcn5 and FeeM. (**a**) Superposition of ApxC (*white*) with Gcn5 (*wheat*). The two Gcn5 substrates are CoA (*red*) and histone H3 peptide (*blue*). (**b**) Superposition of ApxC (*white*) with FeeM (*pink*). The FeeM N-lauroyltyrosine ligand is shown in *green*.

Figure 6: Definition of the TAAT active site. (**a**) Location of the postulated substrate-binding groove and close-up of the active site (boxed). Side chains of key active site residues are shown in stick representation. (**b**) TAAT activity of active site variants assayed by *in vivo* toxin activation (erythrocyte lysis, percentage of wild type activity ± SD with immunoblot showing expression underneath. (**c**) Bacterial 2-hybrid experiments testing ApxC dimerization (*left*), interaction with TAAT-binding protoxin peptide (residues 679-736 - HlyA*, *centre*), and ACP binding (*right*) for wild type (WT) ApxC and mutant variants. Interaction was assessed qualitatively by blue colouration and numerically by β-galactosidase activity (right), expressed as a percentage of the wild type (*top row*) ± SD after subtraction of the negative control. (**d**) Solid surface representations of ApxC showing the location of the active site of each monomer in context of the soluble dimer (*blue* lasso), with important active site residues in *yellow*.

Figure 7: A model for ternary complex TAAT function. (**a**) Location of the acyltyrosine ligand (*green*) from the FeeM:acyltyrosine co-structure and histone H3 peptide (*blue*) from the Gcn5:CoA:H3 peptide ternary complex after superposition with ApxC (*white*). (**b**) Proposed locations of the acylACP and protoxin binding sites on opposite faces of the TAAT enzyme (ACP-binding site *green*; protoxin binding site *blue*). The substrates are predicted to meet at the crux of the split β-sheet where the catalytic residues (*yellow*) are located. Arg121 (essential for ACP binding) is annotated.

Figure 1

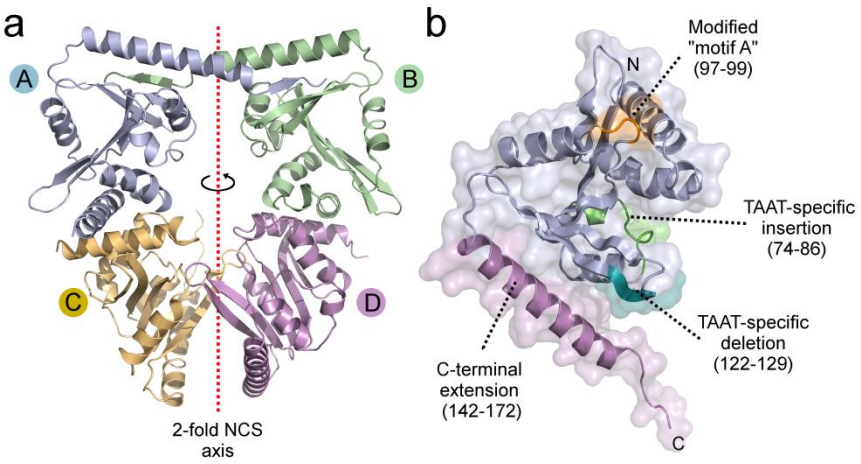

a
A
B
C
D
2-fold NCS
axis
b
Modified
"motif A"
(97-99)
N
TAAT-specific
insertion
(74-86)
TAAT-specific
deletion
(122-129)
C-terminal
extension
(142-172)
C

Figure 2

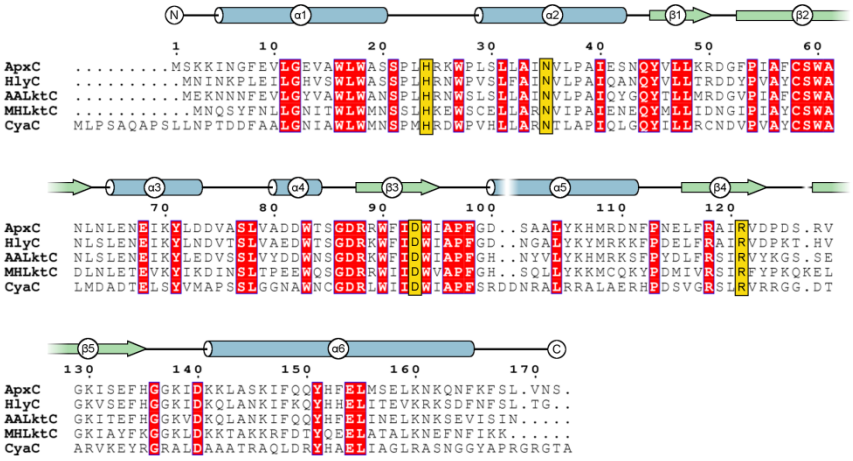

N
α1
α2
β1
β2
1 10 20 30 40 50 60
ApxC .....MSKKINGFEVLGEVAWLWASSPLHRKWPLSLLAINVLPAIESNQYVLLKRDGPIAFCSWA
HlyC .....MNINKPLEILGHVSWLWASSPLHRNWPVSLFAINVLPAIQANQYVLLTRDDYPVAYCSWA
AALktC .....MEKNNNFEVLGYVAWLWANSPLHRNWSLSLLAINVLPAIQYGQYTLLMRDGVPIAFCSWA
MHLktC .....MNQSYFNLLGNITWLWMNSSLHKEWSCELLARNVIPAIENEQYMLLIDNGIPIAYCSWA
CyaC MLPSAQAPSLLNPTDDFAALGNIAWLWMNSPMHRDWPVHLLARNTLAPIQLGQYILLRCNDVPVAYCSWA
α3
α4
β3
α5
β4
70 80 90 100 110 120
ApxC NLNLENEIKYLDDVASLVADDWTSGDRRWFIDWIAPFGD..SAALYKHMRDNEPNELERAIRVDPDS.RV
HlyC NLSLENEIKYLNDVTSLVAEDWTSGDRKWFIDWIAPFGD..NGALYKYMRKKEPDELERAIRVDPKT.HV
AALktC NLSLENEIKYLEDVSSLVYDDWNSGDRKWFIDWIAPFGH..NYVLYKHMRKSEPYDLERSIRVYKGS.SE
MHLktC DLNLETEVKYIKDINSLTPEEWQSGDRRWIIDWVAPFGH..SQLLYKKMCQKYPDMIVRSIRFYPKQKEL
CyaC LMDADTELSYVMAPSSLGGNAWNCGDRLWIIDWIAPFSRDDNRALRRALAERHPDSVGRSLRVRRGG.DT
β5
α6
C
130 140 150 160 170
ApxC GKISEFHGGKIDKKLASKIFQQYHFELMSELKNKQNFKFSL.VNS.
HlyC GKVSEFHGGKIDKQLANKIFKQYHHELITEVKRKSDFNFSL.TG..
AALktC GKITEFHGGKVDKQLANKIFQQYHFELINELKNKSEVISIN.....
MHLktC GKIAYFKGGKLDKKTAKKRFDTYQEELATALKNEFNFIKK.....
CyaC ARVKEYRGRALDAAATRAQLDRYHAELIAGLRASNGGYAPRGRGTA

Figure 3

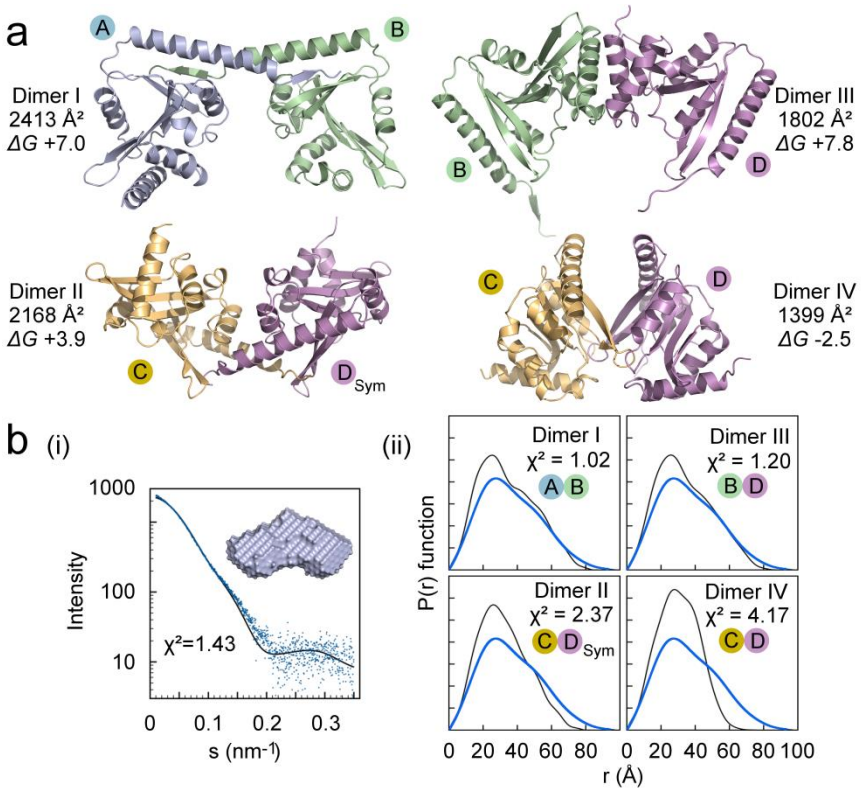

a
A
B
Dimer I
2413 Å²
ΔG +7.0
Dimer III
1802 Å²
ΔG +7.8
B
D
Dimer II
2168 Å²
ΔG +3.9
C
D Sym
C
D
Dimer IV
1399 Å²
ΔG -2.5
b (i)
(ii)
Intensity
1000
100
10
χ²=1.43
0 0.1 0.2 0.3
s (nm⁻¹)
P(r) function
Dimer I
χ² = 1.02
A B
Dimer III
χ² = 1.20
B D
Dimer II
χ² = 2.37
C D Sym
Dimer IV
χ² = 4.17
C D
0 20 40 60 80
0 20 40 60 80 100
r (Å)

Figure 4

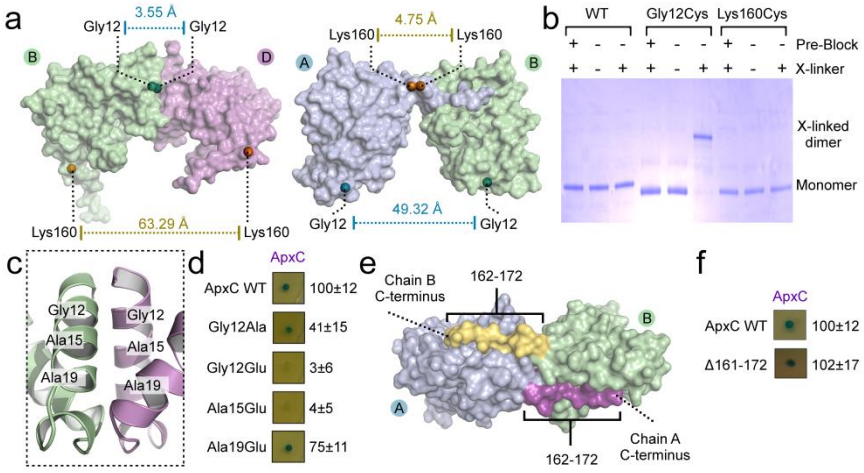

a
3.55 Å
Gly12
Gly12
B
D
Lys160
63.29 Å
Lys160
4.75 Å
Lys160
Lys160
A
B
Gly12
49.32 Å
Gly12
b
WT
Gly12Cys
Lys160Cys
Pre-Block
X-linker
X-linked dimer
Monomer
c
Gly12
Gly12
Ala15
Ala15
Ala19
Ala19
d
ApxC
ApxC WT 100±12
Gly12Ala 41±15
Gly12Glu 3±6
Ala15Glu 4±5
Ala19Glu 75±11
e
Chain B C-terminus
162-172
B
A
Chain A C-terminus
162-172
f
ApxC
ApxC WT 100±12
Δ161-172 102±17

Figure 5

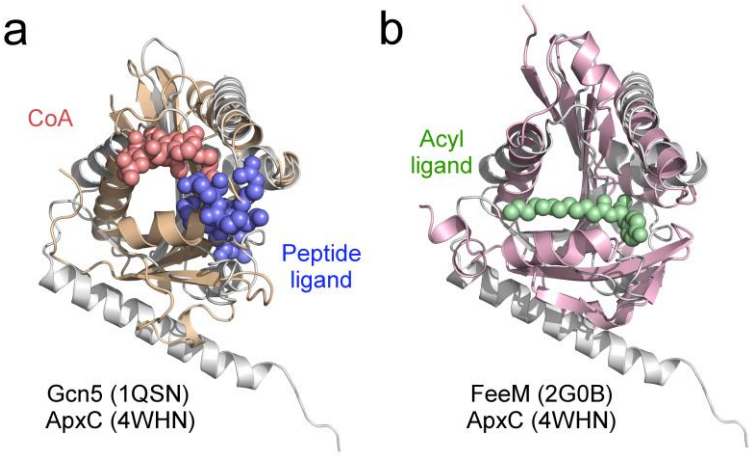

a
CoA
Peptide ligand
Gcn5 (1QSN)
ApxC (4WHN)
b
Acyl ligand
FeeM (2G0B)
ApxC (4WHN)

Figure 6

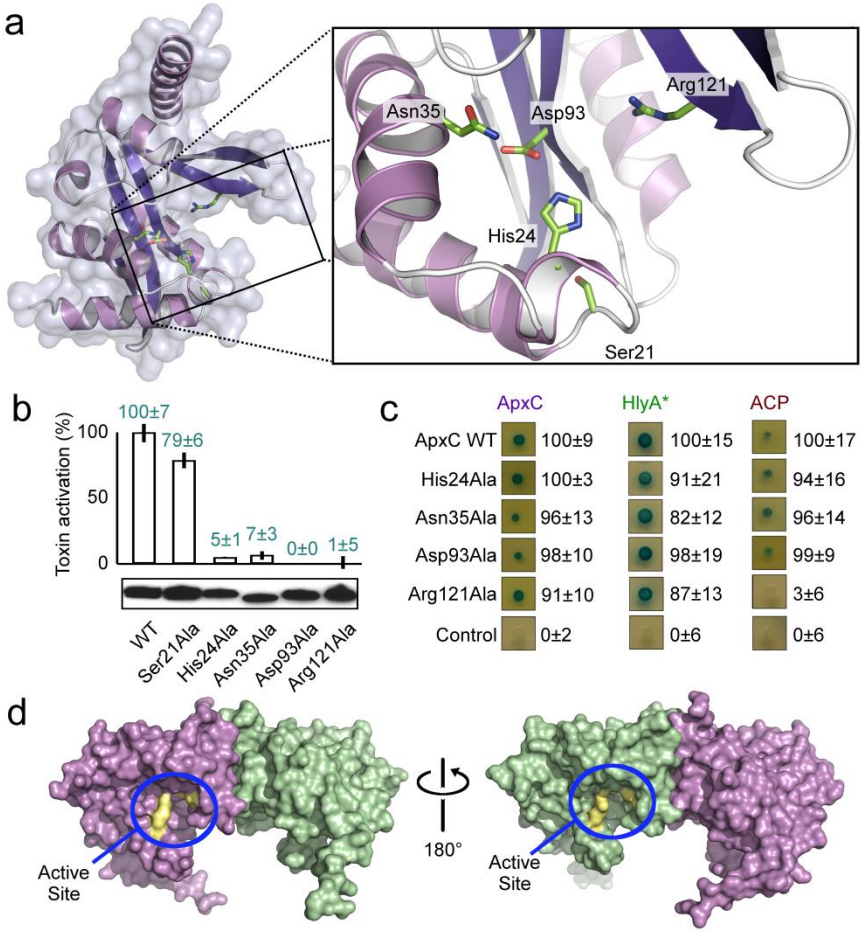
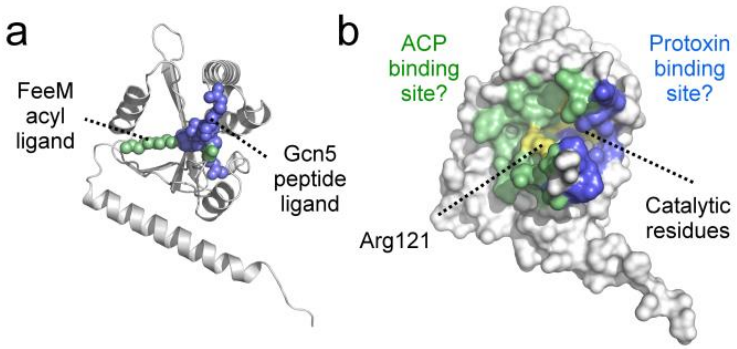


Figure 7


a
FeeM
acyl
ligand
Gcn5
peptide
ligand
b
ACP
binding
site?
Protoxin
binding
site?
Catalytic
residues
Arg121

SUPPLEMENTAL INFORMATION

Structure of a bacterial toxin-activating acyltransferase

Nicholas P Greene, Allister Crow*, Colin Hughes[†] and Vassilis Koronakis*

*Equal contributions.

[†]Corresponding Author.

Supplemental table 1: Crystallographic data and refinement statistics

	Se-Methionine Derivative	High Resolution Native
Data Collection		
Beam line	DLS I24	DLS I04
Wavelength (Å)	0.9700	0.9796
Crystal Parameters		
Space Group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit Cell Dimensions (Å)	73.8, 88.8, 130.8	82.45, 86.37, 131.16
Unit Cell Angles (°)	90, 90, 90	90, 90, 90
Mosaic Spread (°)	0.87	0.37
Reflection Data		
Resolution Range (Å)	88.78-2.65 (2.79-2.65)	54.29-2.15 (2.22-2.15)
Unique Reflections	25,047 (3,129)	51,711 (4442)
R_{sym}	0.189 (0.679)	0.104 (0.804)
I/σ(I)	11.7 (2.9)	12.6 (2.5)
Completeness (%)	97.6 (86.1)	100 (100)
Anomalous Completeness (%)	97.2 (84.0)	100 (100)
Multiplicity	10.8 (7.0)	7.2 (7.3)
Anomalous Multiplicity	5.7 (3.6)	3.7 (3.7)
Wilson B ($Å^2$)	43.8	25.9
Phase Calculations		
Phasing Method	SAD	Molecular replacement with SeMet model
Figure of Merit	0.24	-
Figure of Merit (after density modification)	0.65	-
Refinement †		
Resolution (Å)	-	54.29-2.15
Number of Reflections	-	49,014
R_{work}	-	0.200
R_{free}	-	0.242
R_{cryst}	-	0.202
Rms(Bond Lengths) (Å)	-	0.016
Rms(Bond Angles) (°)	-	1.729
Model Composition		
Protein atoms	-	5,636
Waters	-	219
Other	-	13
Model B-factors		
Protein atoms ($Å^2$)	-	39.7
Waters ($Å^2$)	-	36.0
Other ($Å^2$)	-	65.1
Ramachandran Statistics ‡		
Favoured (%)	-	99.7
Allowed (%)	-	0.3
Outliers (%)	-	0

† *As reported by Refmac.*
‡ *As reported by Rampage.*

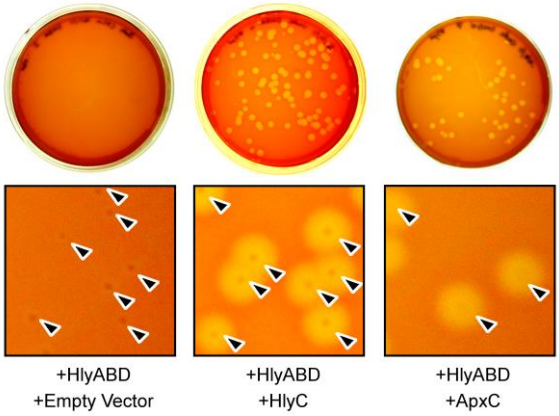
Supplemental table 2: Superposition statistics for matching ApxC to key GNATs.

PDB code	Protein	RMSD	Dali Z-Score	Putative GNAT active site residues	Spatial equivalents in TAAT, ApxC (4WHN)
2REE	CurA	2.8Å	8.6	-	-
1KUV	Serotonin acyl transferase	3.0Å	8.2	-	-
1BOB	Hat1	4.0Å	2.9	Glu255, Asp256	Ala119, Ile120
1CM0	P300/CBP	3.1Å	7.9	Glu570, Phe617, Tyr616	Trp90, *absent*, *absent*
1YGH	Gcn5 (yeast)	3.1Å	7.3	Glu173	Trp90
1QSN	Gcn5 (*Tetrahymena*)	3.2Å	7.8	Glu122	Trp90
2G0B	FeeM	3.6Å	6.6	Glu94	Trp90
1MJA	Myst	4.0Å	3.1	Cys304, Glu337	Asp93, Ala119

Supplemental Table 3: Primers used in this study.

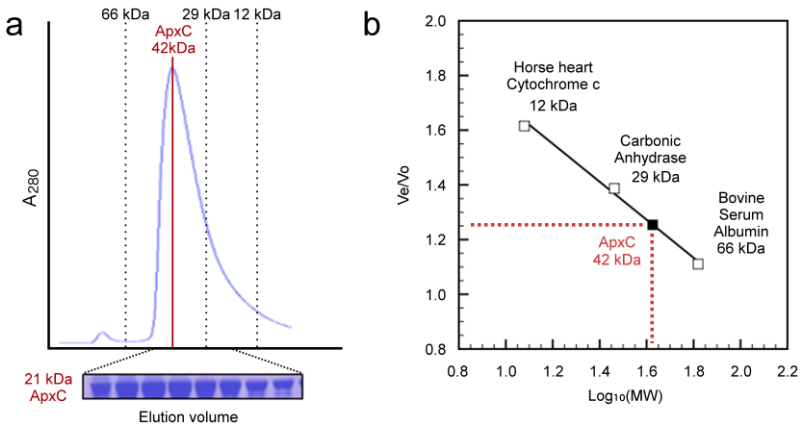
Primer	Sequence 5′→3′
ApxC QE80 F	GCGCGGATCCAGTAAAAAAATTAATGGATTTGAGGTTTTAG
ApxC QE80 R	GCGCAAGCTTTTAGCTATTTACTAATGAAAATTTAAAATTTTG
ApxC HindIII F	GCGCAAGCTTGATGAGTAAAAAAATTAATGGATTTGAG
ApxC SacI R	GCGCGAGCTCCCGCTATTTACTAATGAAAATTTAAAATTTTG
ApxC G12A F	GGATTTGAGGTTTTAGCAGAGGTGGCATGGTTATG
ApxC G12A R	CATAACCATGCCACCTCTGCTAAAACCTCAAATCC
ApxC G12E F	GGATTTGAGGTTTTAGAAGAGGTGGCATGGTTATG
ApxC G12E R	CATAACCATGCCACCTCTTCTAAAACCTCAAATCC
ApxC A15E F	GGTTTTAGGAGAGGTGGAATGGTTATGGGCAAG
ApxC A15E R	CTTGCCCATAACCATTCCACCTCTCCTAAAACC
ApxC A19E F	GGTGGCATGGTTATGGGAAAGTTCTCCTTTACATC
ApxC A19E R	GATGTAAAGGAGAACTTTCCCATAACCATGCCACC
ApxC S21A F	GCATGGTTATGGGCAAGTGCTCCTTTACATCGAAAG
ApxC S21A R	CTTTCGATGTAAAGGAGCACTTGCCCATAACCATGC
ApxC H24A F	GTTATGGGCAAGTTCTCCTTTAGCTCGAAAGTGGCCG
ApxIX H24A R	CGGCCACTTTCGAGCTAAAGGAGAACTTGCCCATAAC
ApxC N35A F	GTGGCCGCTTTCTTTGTTAGCAATTGCTGTGCTACCTGCG
ApxC N35A R	CGCAGGTAGCACAGCAATTGCTAACAAAGAAAGCGGCCAC
ApxC C58A F	CGTGACGGTTTTCCTATTGCATTTGCTAGCTGGGCAAATTTG
ApxC C58A R	CAAATTTGCCCAGCTAGCAAATGCAATAGGAAAACCGTCACG
ApxC D93A F	CGTCGATGGTTTATAGCTTGGATAGCACCGTTCG
ApxC D93A R	CGAACGGTGCTATCCAAGCTATAAACCATCGACG
ApxC R121A F	CTGTTTAGGGCTATTGCAGTTGATCCGGACTCTC
ApxC R121A R	GAGAGTCCGGATCAACTGCAATAGCCCTAAACAG
ApxC 161 R	GCGCGAGCTCCCATTTTTTAGCTCACTCATTAATTCAAAG
HlyA 679 F	GCGCAAGCTTGATGGAAGTTGTGAAGGAGCAGGAGGTTTC
HlyA 736 R	GCGCGAGCTCCCATCAGTAAATTTACTGCCAAAAAACTTG
ACP F	GCGCAAGCTTGATGAGCACTATCGAAGAACGCGTTAAG
ACP R	GCGCGAGCTCCCCGCCTGGTGGCCGTTGATGTAATC

Restriction sites underlined

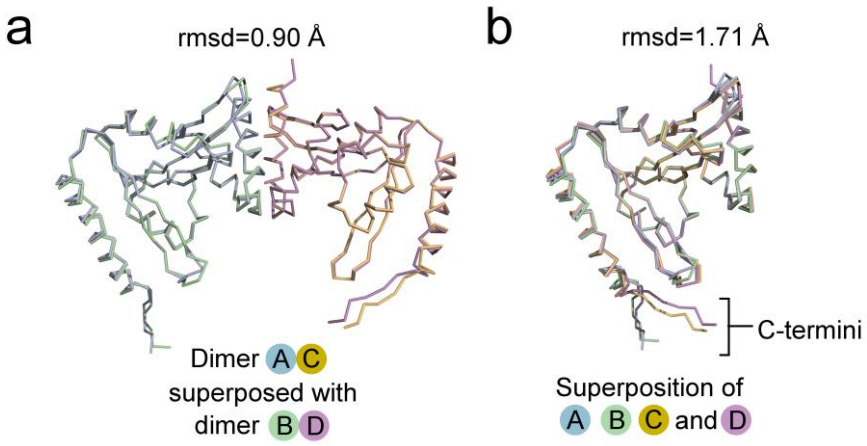


Supplemental Figure 1: ApxC substitutes for HlyC in activating the *E.coli* hemolysin protoxin.

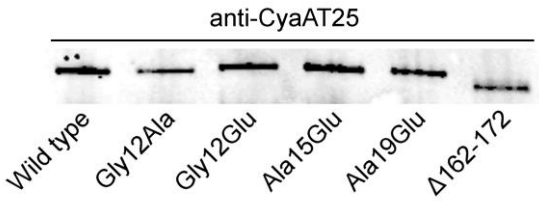
E.coli DH5α (TolC positive) transformed with plasmids encoding the hemolysin protoxin (LG583-HlyA), export proteins (LG575-HlyBD) and either pQE80 empty (control *left*), or encoding a TAAT (HlyC *centre;* ApxC *right*) and grown on blood agar plates. In the absence of a TAAT, colonies (arrowed) are non-hemolytic, but expression of either *E.coli* HlyC or ApxC restores the hemolytic phenotype. Erythrocyte lysis assays confirmed HlyC and ApxC TAAT activities are comparable (100±5 % and 99±6 % activity respectively).



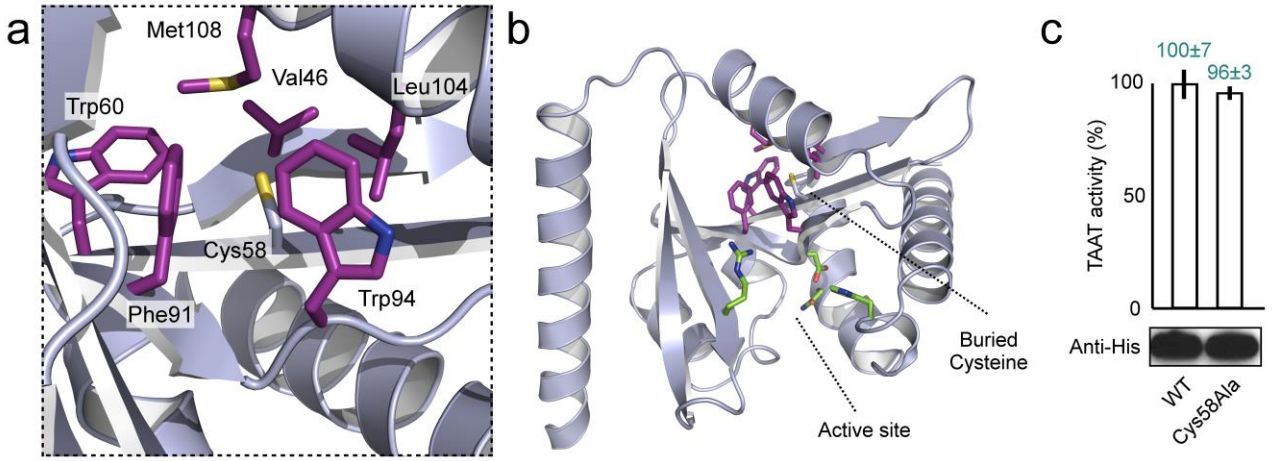
Supplemental Figure 2: Gel filtration of natively-purified recombinant ApxC. (**a**) The A280 trace is shown in *blue* and the elution peak marked in *red,* with SDS PAGE of fractions immediately below. Dotted lines denote the elution maxima of protein standards bovine serum albumin 66kDa, carbonic anhydrase 29kDa, and cytochrome *c* 12kDa. (**b**) Standard linear calibration plot of Ve/Vo (elution volume (mL) divided by void volume of the column (mL)) verses the base-10 logarithm of the molecular weight (kDa) for the three protein standards (white squares). A linear fit to the protein standards is shown as a solid black line (R^2=0.99). The point at which the measured Ve/Vo value for ApxC intersects the calibration curve is denoted by a black square with the measured elution point and corresponding log-molecular weight for ApxC shown as red dashed lines. From the calibration curve, ApxC elutes at an apparent molecular weight of 42 kDa consistent with a dimer of two 21 kDa monomers.



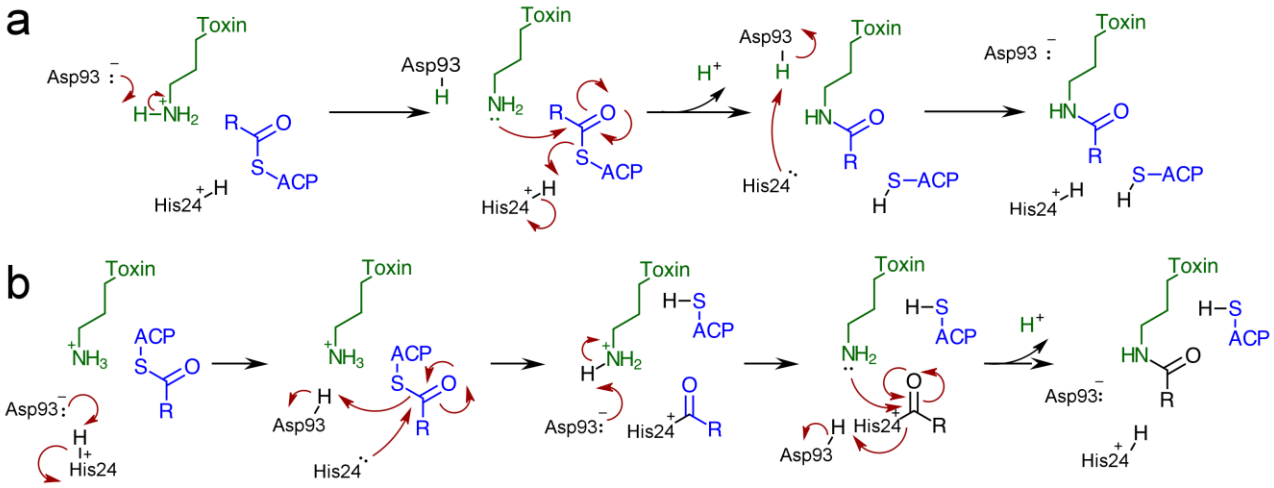
Supplemental Figure 3: Superposition of a pair of near-identical dimers related by 2-fold non-crystallographic symmetry. (**a**) Superposition of a pair of dimers derived from the asymmetric unit; the first composed of monomers A and C, the second of monomers B and D. The root mean square deviation (rmsd) for all Cα positions is 0.90 Å indicating the two arrangements are essentially identical. (**b**) Superposition of all 4 monomers. Rmsd indicated is the mean value for all possible pairwise superpositions of monomers. Deviations in structure are largely limited to the extreme N- and C-termini.



Supplemental Figure 4: Western blot of ApxC-CyaA T25 domain fusions. ApxC-CyaA T25 domain fusions used in the two-hybrid assays (**Fig. 4d**) were expressed in *E.coli* MC1061, and monitored by immunoblot using rabbit anti-CyaA primary (Santa Cruz Biotech) and donkey anti-rabbit dye-conjugated secondary (Licor) antibodies.



Supplemental Figure 5: The conserved TAAT cysteine is located in the protein core and is not required for TAAT activity. (**a**) Close-up of the conserved cysteine. (**b**) Context of Cys58 within the ApxC monomer. Cys58 is located within the hydrophobic core of the protein on the second β-strand. The side chain thiol (*yellow*) is located on the opposite side of the central sheet to residues that belong to the active site (*green*). Hydrophobic residues (*purple*) pack around the cysteine, sealing-in the side chain and preventing access to potential substrates. The location of the conserved cysteine precludes the possibility of a catalytic role for this residue. (**c**) TAAT activity of wild type (WT) and Cys58Ala ApxC (Western blot shows expression) assayed by *in vivo* activation of HlyA protoxin, read by erythrocyte lysis. Values above the bars indicate percentage of wild type activity ±SD.



Supplemental Figure 6: Potential mechanisms for toxin activation by TAATs. (**a**) The direct attack mechanism. (**b**) The covalent catalysis mechanism. Enzymatic residues are *black*, the target toxin lysine is *green* and acyl-ACP *blue*. The acyl chain is denoted 'R'. Curly arrows (*red*) denote electron pair movements. Possible roles for essential residues Asp93 and His24 are shown in context of each mechanistic scheme.