The three-dimensional structure of a superantigen-like protein, SET3, from a pathogenicity island of the *Staphylococcus aureus* genome

Vickery L. Arcus¹, Ries Langley², Thomas Proft², John D. Fraser² and Edward N. Baker¹*

¹School of Biological Sciences and ²Department of Molecular Medicine, University of Auckland, Private Bag 92-019 Auckland, New Zealand.

Running title: Three-dimensional structure of SET3

*Corresponding author:
Edward N. Baker
School of Biological Sciences
University of Auckland
Private Bag 92-019
Auckland
New Zealand
Ph: 64 9 373 7599 ext 4415
Fax: 64 9 373 7619
e-mail: ted.baker@auckland.ac.nz
SUMMARY

The Staphylococcal Enterotoxin-like Toxins (SETs) are a family of proteins encoded within the *Staphylococcus aureus* genome that were identified by their similarity to the well-described bacterial superantigens. The first crystal structure of a member of the SET family, SET3, has been determined to 1.9 Å ($R=0.205$, $R_{free}=0.240$) and reveals a fold characteristic of the superantigen family but with significant differences. The SET proteins are secreted at varying levels by staphylococcal isolates and sero-conversion studies of normal individuals indicate that they are strongly antigenic to humans. Recombinant SETs do not exhibit any of the properties expected of superantigens such as major histocompatability complex class II (MHC-II) binding, or broad T-cell activation, suggesting they have an entirely different function. The fact that the whole gene family is clustered within the pathogenicity island SaIn2 of the *S. aureus* genome suggests that they are involved in host/pathogen interactions.
INTRODUCTION

The bacterial superantigen (SAg) family is a large protein family exclusive to three pathogenic species: *Staphylococcus aureus*, *Streptococcus pyogenes* and *Streptococcus equi*. The former two organisms are opportunistic human pathogens and the latter is a pathogen of horses. Members of this family of proteins have been implicated in a range of human diseases, including staphylococcal food poisoning, scarlet fever, toxic shock, rheumatoid arthritis and secondary HIV infection (1). *Streptococcus equi* is the causative agent of strangles in horses (2). Superantigens function by immune modulation. They cross-link major histocompatibility complex class II (MHC-II) and T-cell receptor (TCR) molecules, causing non-specific and disproportionate T-cell proliferation and cytokine release (3). This gives rise to symptoms characteristic of fever and toxic shock. As a result of their close association with serious human pathologies, superantigens have been the subject of intense research over the last decade and several reviews have been published (3-5).

The identification of a Staphylococcal gene cluster (6) and the determination of the complete genome sequence of *S. aureus* (7) have revealed a family of genes with similarity to superantigens from *Staphylococcus aureus*. These were first described by Williams and colleagues (6), as a cluster of five related genes whose protein products were reported to stimulate the production of IL-1β, IL-6 and TNF-α from human peripheral blood mononuclear cells. The complete genome sequences for two strains of *S. aureus* subsequently showed that this cluster contains a total of ten genes in one strain and nine in the second strain, with these genes being related by sequence identities of 36-67%. The SET gene cluster represents half of a pathogenicity island, SaPIn2, which is shown in Figure 1. Downstream (3’) to the SET gene cluster is a set of 9 *lpl* genes which contain lipoprotein attachment sites and are thought to code for pathogenic proteins (7). The pathogenicity island
is flanked at the 5’ end by a putative transposase and is also thought to be maintained in the genome by two genes, \textit{hsdS} and \textit{hsdM}, which are part of a three component restriction-modification system.

The SET genes are most closely related, by sequence, to toxic shock syndrome toxin (TSST), which is a Staphylococcal superantigen. The overall sequence identity is low, at only 26%, but there are small regions which are very well conserved when compared to superantigens. Thus, the SET proteins carry the SAg signature sequences. Previous structural and functional studies of superantigens have revealed the capacity for these proteins to graft a wide variety of functional modes onto a conserved fold (5,8). For example, different superantigens variously bind to MHC-II molecules \textit{via} their N-terminal domain (SEB, TSST) (9,10) or \textit{via} a zinc atom bound on the surface of their C-terminal domain (SMEZ-2, SPE-C) (11). They are able to cross-link MHC-II molecules at the cell surface by simultaneous N- and C-terminal binding in the case of SEA (12) or by dimerisation in the case of SPE-C (13). The superantigens also show allelic variation amongst strains (SMEZ-2) (14) and varying degrees of specificity for alleles of the TCR (15).

The availability of genomic sequence data has transformed our knowledge of the distribution of superantigens and superantigen-like proteins in bacterial species. The completion of the \textit{Streptococcus pyogenes} genome sequence (16) revealed 6 new superantigens whose structures and functions have been characterised (17). Although it is unsurprising that the Staphylococcal genome would also encode a new group of superantigen-like proteins, the structural and functional variation in this family leaves many questions as to their activity. Here, we describe the first structure of a protein from the SET gene cluster, SET3. The structure reveals that although the superantigen fold is conserved and by inference, is also
conserved among the other SET proteins, there is evidence that they do not function as prototypic superantigens. We also show, however, that they are likely to have a significant role in host/pathogen interactions in humans, because seroconversion is widespread and antibody titres in human serum are high in the majority of tested cases.

**EXPERIMENTAL PROCEDURES**

*SET3 protein expression and purification.* The *set3* gene was amplified by PCR from genomic DNA isolated from a local hospital strain (Auckland, New Zealand) of *Staphylococcus aureus*. The gene was sub-cloned into the expression vector pET32a_3C and expressed in *E. coli* (AD494) as an N-terminal thioredoxin fusion protein. The fusion protein was purified by Ni²⁺ affinity chromatography, cleaved using 3C protease, and subjected to Ni²⁺ affinity chromatography again to separate SET3 from thioredoxin. Pure SET3 was finally obtained by size exclusion chromatography (Superdex75, Pharmacia). Dynamic light scattering data for the final concentrated protein solution (12 mg.ml⁻¹ SET3) indicated a monodisperse solution of monomeric protein. The molecular weight calculated from the hydrodynamic radius was 29 kDa, in comparison to the calculated molecular weight of 24.16 kDa. Other SET proteins for biochemical studies were produced similarly.

*¹²⁵I labelling of proteins.* SET3 (6µg) was mixed with 0.5 mCi ¹²⁵I in 20 µl of 5xPBS and 5 µl 1 mg.ml⁻¹ Chloramine T, and incubated for 1 min. 5 µl of 5 mg.ml⁻¹ sodium metabisulphite was added to this solution and the volume made up to 50 µl with PBS/1% FCS. This was then loaded on to a 1 ml G25 column pre-equilibrated with PBS/1% FCS and the protein was eluted in 100 µl fractions. An aliquot from each fraction (1 µl) was counted using a Cobra II gamma counter.
**Immunoprecipitation to detect for seroconversion.** 5 µl of human serum was incubated with 10^5 cpm SET3^{125}I in a total volume of 50 µl TSA/10%FCS/1% Hb at 37 °C for 30 min. 5 µl of protein A-sepharose was added and the sample incubated on ice for 30 min. After washing 3 times in 0.5 ml RIPA buffer, the sample was analysed using a Cobra II gamma counter. A control containing no serum was used to determine the background count.

**Western Analysis of S. aureus culture supernatants.** *Staphylococcus aureus* clinical isolates were cultured overnight without shaking at 37°C. After centrifugation to remove the bacteria, 0.5 ml of culture supernatant was concentrated by precipitation with 50% TCA and resuspended in 5 µl PBS. Samples were subjected to electrophoresis on a 12% SDS-PAGE protein gel. The proteins were transferred to nitrocellulose using a semi-dry blotter. The membrane was blocked for 1 hour at room temperature in TTBS/5% (w/v) non-fat milk powder. A 1/500 dilution of mouse anti-SET3 polyclonal serum was then incubated with the membrane for 1 hour at room temperature in blocking buffer. After washing in TTBS the membrane was incubated with a 1/1000 dilution of peroxidase labelled anti-mouse IgG for 1 hour at room temperature. The membrane was then washed in TBS and analysed using an ECL Western blotting detection kit from Amersham Pharmacia Biotech.

**PBL proliferation assay.** A five fold serial dilution (in triplicate) from 1 µg.ml^{-1} to 2.05x10^{-8} µg.ml^{-1} of each protein tested was set up in 96-well plates in a volume of 100 µl RPMI/10% FCS per well. One row of wells was set up with no protein added as a negative control. Mononuclear cells were isolated from whole blood by Ficoll separation, washed and suspended at 10^6 cells/ml in RPMI/10% FCS. 100 µl of cells were added to each well and the plates were incubated in 5% CO₂ for 3 days. 25 µl of [³H]thymidine was added to each well...
and incubated overnight. The cells were harvested onto filter mats using a Tomtec harvester and the incorporated $^3$H-TdR was analysed using a trilux counter.

**SET3 crystallisation.** SET3 crystals were grown in hanging drops at 18 °C by mixing 2 µl of protein solution (20 mM sodium phosphate buffer, pH=6.0, 12 mg.ml$^{-1}$ SET3) with 2 µl of precipitant solution (either 50 mM HEPES, pH=7.0, 1.5 M ammonium sulphate or 50 mM HEPES pH=7.0, 0.9 M Na/K tartrate). Hexagonal plates and blocks typically emerged after four days and then grew larger over two weeks. X-ray diffraction measurements showed that the crystals were hexagonal, space group $P6_3$, with unit cell dimensions $a=b=65.1$, $c=196.7$ Å. Two molecules occupy the asymmetric unit, giving a solvent content in the crystal of 45% (v/v). The crystals could be flash-frozen for data collection by soaking in cryoprotectant (mother liquor plus 30% glycerol) immediately prior to placement in a stream of cold N$_2$ gas (110 K).

**X-ray data collection.** Native and derivative data sets for SET3 were collected using CuKα radiation from a Rigaku RU-H3R X-ray generator equipped with focussing mirrors and a Mar345 imaging plate detector. Subsequently, a high resolution native data set was collected using synchrotron radiation ($\lambda = 0.8452$ Å) at DESY Hamburg, beamline BW7V. The raw data were processed using DENZO and subsequently scaled using SCALEPACK (18). Data collection statistics are given in Table 1.

**Structure determination and refinement.** The structure of SET3 was determined by multiple isomorphous replacement (MIR), using three heavy atom derivatives. These were prepared by soaking SET3 crystals in KAu(CN)$_2$ (5 mM), K$_2$PtCl$_6$ (1 mM) and Hg(CH$_3$COO)$_2$ (1 mM) for periods of 1-7 days. The MIR analysis was carried out in the higher symmetry space group $P6_3$.
group of \(P6_{3}22\) (one molecule in the asymmetric unit), using the program SOLVE (19), incorporating isomorphous and anomalous differences. Two sites were found for the KAu(CN)\(_2\) derivative and single, identical, sites were found for the K\(_2\)PtCl\(_6\) and Hg(CH\(_3\)COO)\(_2\) derivatives. Initial phases from SOLVE were calculated to 2.8 Å and gave a figure of merit of 39% and a z-score of 16.7. These phases were checked by refinement using MLPHARE (CCP4 program suite, 1994), extended to 2.3 Å and then improved by maximum likelihood density modification using RESOLVE (20). The program MAID (21) was used to auto-trace 65 residues, and subsequently a further 44 residues were manually built using O (22). Side chains were added for 23 residues, with the remainder left as alanine. A single round of phased refinement indicated that the symmetry constraints of \(P6_{3}22\) were too high and the space group symmetry was therefore lowered to \(P6_{3}\) with two molecules in the asymmetric unit. The model, now comprising 218 residues, was used to determine initial phases for WARP (23) and the remainder of the two molecules were built automatically by WARP using data in the resolution range 30 - 1.90 Å. Finally, five iterations of manual building and refinement using O and CNS (24) were undertaken to finish the structure determination. Final statistics are given in Table 2. Structural comparisons were performed using the program MAPS (http://bioinfo1.mbfys.lu.se/TOP/maps.html; Lu, 1998) and comparison trees were constructed from structure diversity scores as previously described (8).
RESULTS

Quality of the Structural Model. The three-dimensional structure of SET3 was determined by X-ray crystallography and refined using data to 1.90 Å resolution. The final model comprises residues 5–204 of both molecule 1 and molecule 2, these comprising the two molecules in the crystal asymmetric unit, together with 378 solvent molecules, modelled as water. Three or four residues (83–86 molecule A and 84–86, molecule B) from the β4–β5 loop in the N-terminal domain, have poorly defined density in each molecule and have not been modelled; these residues are assumed to be disordered.

Overall, the fit to the electron density is excellent and the structure conforms well with the X-ray data and with expected protein geometry. The final \( R \) (\( R_{\text{free}} \)) values are 0.205 (0.240) and 90.6% of the 393 residues in the two independent molecules have \( \phi/\psi \) torsion angles in the most favoured regions of the Ramachandran plot, as defined by PROCHECK (CCP4, 1994). Two residues in each molecule (Asp104 and Lys160, with \( \phi/\psi \) angles (-142,-115) and (48,-122) respectively) are in generously allowed regions of the plot, and one residue in each molecule (Asp118, \( \phi/\psi \) angles (49,-112)) occupies a disallowed region. Both Asp118 and Lys60 are central residues in type II’ β-turns and this accounts for their unusual \( \phi/\psi \) angles (see Figure 2). Asp118 is part of the tight turn that links β-strands β6 and β7. However, Lys60 is part of a β-turn which does not link secondary structure elements and is flanked on both sides by loops. Asp104 lies between two tyrosine residues in a tightly-packed structure that includes one leucine and three tyrosine residues, which may constrain Asp104 in its unusual \( \phi/\psi \) geometry. This Asp residue is conserved in 6 out of 10 SET proteins. The electron density for each of these three residues (Asp118, Lys60, Asp104) in each SET molecule in the asymmetric unit is unequivocal and, at least in the case of Lys60 and Asp104,

Arcus et al.
their consistent, non-standard geometry points towards their possible significance as functional residues for SET3.

*Structure description.* SET3 has the classic two-domain structure that is characteristic of proteins of the superantigen family (Figure 2). Most of the N-terminal half of the molecule (residues 24-97) is folded into a 5-stranded β-barrel called an OB-fold (25), and the C-terminal half, residues 98-204, into a β-grasp domain, in which a 5-stranded mixed β-sheet packs over an amphipathic α-helix that is the most sequence-conserved feature of these proteins. The two domains are bridged by an N-terminal helix, residues 11-23. The two domains pack together to form an irregularly shaped molecule with approximate dimensions 50x37x23 Å³.

The SET3 structure has two notable features that differentiate it from other SAg-family proteins whose structures have been solved to date. First, the strands β6 and β7 in the C-terminal domain are extended so that the β6-β7 loop projects from the molecular surface (the strands flank D118, Figure 2). In the crystal, the two independent molecules in the asymmetric unit are related by a pseudo two-fold axis about β7, and this extended loop forms a β-sheet with the same loop of the second molecule. The surface area buried in the crystallographic dimer interaction is small (287 Å² per monomer) and it appears that this “dimer” does not persist in solution, since dynamic light scattering indicates a monomer. However, the formation of this dimer includes 6 intermolecular hydrogen bonds and constructs a large and continuous β-sheet of 10 strands between the two monomers. This, in turn, presents a large, positively charged, saddle-shaped surface (Figure 2) that has the potential to act as a binding surface for negatively charged molecules such as DNA. This highlights a second striking feature of the structure, in the widespread positive charge over...
the protein surface. Although lysine and arginine residues are scattered across the surface they are concentrated at the capping helix and outer face of the N-terminal OB-fold domain, and in the saddle-shaped face formed by the association of the two molecules in the asymmetric unit described above. We suggest that these areas are likely binding sites for negatively charged binding partners.

Comparison with other superantigens. Proteins from the superantigen family are well represented in the current structural database, SET3 being the fourteenth such structure from this family to be determined. When these proteins are compared as a whole, it is apparent that the structural core is conserved, with 76 residues occupying equivalent Cα positions, with a root-mean-square (rms) difference of 1.15 Å (for those 76 residues), but with only five residues conserved at the sequence level. SET3 and TSST form an out-group from the remainder of the SAg-like proteins, such that if these two structures are omitted from the alignment, 113 residues share equivalent Cα positions (with an rms difference of 0.95 Å for those 113 residues) and 16 residues are conserved in the sequence. These 16 residues appear to be conserved for their role in preserving the superantigen fold, and apart from glycines are almost exclusively involved in buried charge-charge or polar interactions between elements of secondary structure. Most are concentrated around the interdomain region and appear to help maintain the close association of the two domains.

When the SET3 and TSST structures are overlaid, 127 Cα positions are within 3.5 Å with an rms difference of 1.15 Å, with 37 sequence identities in this conserved core. Again, the majority of the residues that are conserved in sequence and in structure between TSST and SET3 play clear structural roles either contributing to the hydrophobic core, or with their side chains forming buried charge-charge interactions or hydrogen bonds with main-chain atoms.
The exceptions to this are Lys114, Lys160 and Glu200 (SET3 numbering), all on the outer face of the C-terminal β-grasp domain, and Asn150, Asp16 and Asp19, on the opposite face of the β-grasp domain, in proximity to the N-terminal helix. These six residues do not, however, correspond to any equivalent residues in TSST that are involved in either MHC-II binding or TCR interactions (10). In this context, the structure suggests that it is most unlikely that SET3 could bind to either MHC-II or the TCR in a manner similar to TSST. For example, 15 residues from TSST interact with MHC-II upon binding (10) and, after structural alignment, just 2 of these residues are conserved between TSST and SET3 (see Figure 3). Of the remaining 13 MHC-II binding residues, the differences between TSST and SET3 include Asp→Tyr, Arg→Ile, Leu→Arg and Pro→Gly. These non-conservative changes at the binding face argue strongly against any similar binding of SET3 to MHC II.

There are other significant structural differences between TSST and SET3. In comparison to TSST, SET3 has an elongated loop (residues 37-42) that connects strands β1 and β2 on the surface of its N-terminal domain. In SET3, the N-terminal β-barrel is also capped by an α-helix (residues 58-68, linking β3 and β4) that is missing in TSST but present in all the other superantigens and is also a general feature of the OB-fold (25, Figure 3). The loop that leads into the C-terminal domain, residues 99-107, has a very different conformation in SET3 when compared with TSST. In SET3 this loop is solvent-exposed and contains Asp104, which occupies a generously allowed region of the Ramachandran plot and may be important for SET3 function. As described above, SET3 also has elongated β-strands, β6 and β7, at the periphery of the C-terminal β-sheet (residues 116-129).

Allelic variation. There is preliminary evidence that SET3 has allelic variants in different strains of *Staphylococcus aureus*. Two strains have been sequenced from isolates from...
Greenlane Hospital in Auckland, New Zealand, and these are both different from each other and from the two sequences from Williams et al. 2000, and Kuroda et al. 2001. When these four SET3 proteins are aligned they vary at 35 positions. 17 of these positions are concentrated on the two outer β-strands of the C-terminal β-sheet, β6 and β7, whereas the remaining 18 positions are scattered across the protein (see Figure 4). Five of the 37 residues are buried hydrophobic residues and their allelic variants are also hydrophobic, but all other residues are surface exposed. Allelic variation has also been reported for SET1 and is suggested by the sequence differences between the allelic pairs SET1/11, SET2/8, SET3/10, SET4/14 and SET5/13.

The SET family of proteins are not superantigens. Three recombinant proteins from the SET family, SET1, SET3 and SET15, were used in a typical peripheral blood lymphocyte (PBL) stimulation assay in order to assess their mitogenicity. The ability to stimulate the proliferation of T-cells in a TCR Vβ-restricted manner is a hallmark of superantigens. TSST was used as a positive control, and reaches maximum stimulation at 1 ng.ml\(^{-1}\). In all cases and at all concentrations, however, the SET proteins showed no capacity to induce T-cell proliferation (see Figure 5).

SET3 is secreted by all Staphylococcus aureus isolates. Polyclonal antibodies to the SET proteins were raised in mice. α–SET3 was used to detect for the presence of SET3 in the culture supernatants of six S. aureus clinical isolates. These isolates were randomly chosen from a panel of 20 clinical isolates, all of which were found to be positive for the SET3 gene by PCR. Western analysis of the supernatants, concentrated 50 times by TCA precipitation, shows that SET3 is secreted by S. aureus in all cases (see Figure 6).
SET3 is recognised by the human immune system. Radioactively labelled SET1 and SET3 were immunoprecipitated using a panel of human sera together with protein A-sepharose. Serum from 15 healthy individuals was incubated with radioactively labelled SETs to detect for the presence of antibodies against these proteins. All individuals showed positive seroconversion indicating that they have all been exposed to these proteins at some time (see Figure 7).

DISCUSSION

The set family of genes encode a new and large family of SAg-like proteins. As a family they are most closely related to TSST by sequence, and the 9–10 set genes lie consecutively on the Staphylococcus aureus genome (7), almost certainly as a result of gene duplications. Divergence of these genes within the genome is marked and, as is characteristic of the whole superantigen family with which they share sequence similarity, is presumed to have given rise to structurally homologous proteins with different, but related, functions.

The structure of SET3, the first from this family, and the first from the SaPIn2 pathogenicity island in the S. aureus genome, shows clearly that this protein, and by implication, the other SET proteins, belongs to the wider superantigen superfamily. The most striking feature of this superfamily is the diversity of functions (and sequences) supported by their highly conserved structural framework (5,8). Sequence conservation is limited to residues that confer structural integrity, particularly in and around the interdomain region, whereas the protein surfaces vary considerably. Variations include changed external loops and metal binding sites, in addition to changes in specific surface residues, and lead to differences in their interactions with MHC-II and TCR molecules. These variations accompany a remarkably narrow genealogy—a single ancestral gene duplicating in just three known
species of Staphylococcus and Streptococcus, despite the potential for further spread through horizontal gene transfer.

The SET proteins mark a further functional divergence. SET1, 3 and 15 do not stimulate T-cells nor do they bind to MHC-II. Thus, they do not function as non-specific activators of the immune system in the way that the prototype superantigens do. Structurally this is consistent with our observation that in SET3, at least, there is no conservation of the residues that are involved in MHC-II or T-cell receptor binding in the other superantigen structures and sequences.

Without exception, the residues that are strictly conserved amongst the SET family are structural. This gives no clue to the function of SET3, but does strongly point to the likelihood of differing functions for proteins within the SET family. We have modelled seven of the SET sequences onto the SET3 structure, and our hypothesis of varied function is supported from inspection of the surface features of the different SET proteins. For example, SET15 has a low predicted pI of 6.5 (c.f. 9.6 for SET3) and this is reflected in the surface charge on the protein which shows a large area of negative charge on one face of the protein and a large area of positive charge on the opposite face.

The failure to stimulate T-cell proliferation in no way excludes the SET proteins from being important virulence factors. They are secreted by the bacteria, both in vitro and in vivo, and seroconversion is present in the majority of patients that we have tested so far, with high antibody titres in these subjects. The pathogenicity island SaPIn2 appears to be important to the organism as it is maintained in the genome by two restriction-modification genes in the centre of the island (the third member of the trio, the restriction enzyme, lies remote from the
pathogenicity island). Remnants of gene duplication and transfer may also be found in the transposase gene at the 5’ end of SaPln2. This further supports the importance and prevalence of these proteins as agents of host/pathogen interactions.

The superantigen family of proteins continue to be surprising. Recently, the complete genome sequences of both *Streptococcus pyogenes* and *Staphylococcus aureus* have led to the identification of 28 superantigens and superantigen-like proteins. In many individual cases, allelic variation between strains augments this number and the sequence variation covers the continuum from individual amino acid substitutions through to <10% sequence identity between members of the family. The structure of SET3 shows that the conserved superantigen architecture can support yet further functional variation and the presence of nine related homologs adjacent in the genome suggests that these proteins are important to the pathogenicity of Staphylococcus.
REFERENCES


   [http://www.pymol.org](http://www.pymol.org)

FOOT NOTES

*This work was supported by a grant from the Health Research Council of New Zealand.

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

1Abbreviations: SAg, superantigen; MHC-II, major histocompatibility complex class II; TCR, T-cell receptor; PBS, phosphate buffered saline (10 mM PO$_4^{3-}$, pH 8.0, 140 mM NaCl, 3 mM KCl); TBS, tris buffered saline (100mM Tris pH 7.5, 0.9% w/v NaCl); FCS, foetal calf serum; RIPA, radio-immunoprecipitation assay buffer (1% Triton X-100, 1 % sodium deoxycholate, 1% bovine hemoglobin, 1 mM iodoacetamide); TSA, tris buffered saline + azide (10mM Tris pH 8.0, 140mM NaCl, 0.025% NaN$_3$); Hb, Haemoglobin; TCA, trichloroacetic acid; TTBS, 0.1% Tween-20 in TBS; HEPES, N-2-hydroxyethylpiperazine-N’-2-ethansulphonic acid.
FIGURE LEGENDS

Figure 1. Pathogenicity Island, SaPIn2.

The pathogenicity island, SaPIn2, from *Staphylococcus aureus*, strain N315, as defined by (7). Open reading frames are shown with arrows and where genes have been annotated their gene name is given. Other open reading frames are named according to their position in the genome. Figures below each set of arrows indicate approximate genome position in kilobases.

Figure 3. Structural comparison of the N-terminal “binding face”.

A structural alignment of the N-terminal binding faces of SET3, TSST and SPE-C. TSST is shown in grey and residues which interact with MHC-II in the TSST/MHC-II complex are identified by small black spheres and labelled using the one-letter amino acid code. Residues in SET3 and SPE-C at equivalent positions are shown in the sequence alignment at the bottom of the figure. The Cα trace for SET3 is shown in red and the Cα trace for SPE-C is shown in blue.

Figure 2. SET3 Structure.

Above: A stereo view of the Cα backbone trace of SET3 coloured from the N-terminus to the C-terminus (blue to red). Selected residues are labelled according to their single-letter amino acid code and their number in the sequence. A loop spanning five residues (83–86) is not visible in the electron density for the SET3 structure and is omitted in this figure. Below Left: Schematic ribbon diagram of SET3 showing the canonical superantigen fold viewed along the pseudo 2-fold axis of the putative dimer. The N- and C-termini are labelled and the N-terminal OB-fold domain is at the top left and bottom right of the structure. This figure was drawn using PyMol (26). Below right: The electrostatic potential at the protein surface...
looking down the pseudo crystallographic 2-fold axis showing the concentrations of positive charge at each end of the individual monomers. This figure was drawn using GRASP (27).

**Figure 4. Sequence and structural alignment in the superantigen family.**

A sequence alignment for nine members of the superantigen family highlighting conserved features at both sequence and structural levels. Residues which are conserved by sequence are highlighted by red (hydrophobic) and blue (hydrophilic) squares. Cysteines are shown in yellow squares. Regions which are structurally conserved (equivalent Cα positions within 3 Å) are boxed in grey. Positions which show allelic variation for SET3 are shown beneath the alignment and secondary structure of SET3 is also shown schematically. Residues which bind zinc at the interface between the SAg and MHC-II are shown in bold.

**Figure 5. T-cell proliferation.**

Stimulation of human T-cells with recombinant *S. aureus* proteins. Peripheral blood lymphocytes isolated from human blood were incubated with varying concentrations of protein. Stimulation of T-cells was measured by the incorporation of [3H]thymidine as counts per minute (cpm).

**Figure 6. *S. aureus* supernatants.**

Western analysis of culture supernatants from 8 *S. aureus* isolates (lanes 3-10) to detect for the presence of SET3. The supernatants were concentrated 50 times and run alongside 5ng (lane 1) and 1ng (lane 2) of recombinant SET3 by 12% SDS-PAGE. SET3 was detected with mouse anti-SET3 antiserum.
**Figure 7. Seroconversion.**

Immunoprecipitation of human sera to detect for antibodies specific to SET1 and SET3.

A. Serum from 5 individuals incubated with $^{125}$I-SET3 was immunoprecipitated with protein A-sepharose. Levels of specific antibody were measured as a percentage of the total cpm of radioactive protein bound. B. Similarly, serum from 15 individuals was surveyed for the presence of SET1 and SET3.

**Table**

Table 1. Data collection, structure solution and refinement statistics.
Figure 1
Arcus et al.
Figure 2
Arcus et al.
Figure 3
Arcus et al.
Figure 4
Arcus et al.
Figure 5
Arcus et al.
Figure 6
Arcus et al.
Figure 7A
Arcus et al.
Figure 7B
Arcus et al.
Table 1. Data collection, structure solution and refinement statistics.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Native</th>
<th>KAu(CN)₂</th>
<th>K₂PtCl₆</th>
<th>Hg(CH₃COO)₂</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Crystal data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P6₃22</td>
<td>P6₃22</td>
<td>P6₃22</td>
<td>P6₃22</td>
</tr>
<tr>
<td>Cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crystal data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P6₃22</td>
<td>P6₃22</td>
<td>P6₃22</td>
<td>P6₃22</td>
</tr>
<tr>
<td>Cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ax lengths (Å)</td>
<td>65.12, 65.12, 196.73</td>
<td>65.10, 65.10, 197.44</td>
<td>64.74, 64.74, 197.16</td>
<td>64.93, 64.93, 196.97</td>
</tr>
<tr>
<td>angles (°)</td>
<td>90 90 120</td>
<td>90 90 120</td>
<td>90 90 120</td>
<td>90 90 120</td>
</tr>
<tr>
<td><strong>B. Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30–1.95 (2.02–1.95)</td>
<td>25–2.70 (2.80–2.70)</td>
<td>25–2.80 (2.90–2.80)</td>
<td>25–3.10 (3.21–3.10)</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>174729</td>
<td>270501</td>
<td>108469</td>
<td>59994</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>18569</td>
<td>7438</td>
<td>6513</td>
<td>4968</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.8 (97.3)</td>
<td>99.9 (99.0)</td>
<td>98.2 (96.6)</td>
<td>99.4 (98.9)</td>
</tr>
<tr>
<td>Mosaicity</td>
<td>0.29</td>
<td>0.42</td>
<td>0.34</td>
<td>0.67</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>10.8 (40.2)</td>
<td>14.2 (53.1)</td>
<td>14.0 (54.8)</td>
<td>17.5 (42.4)</td>
</tr>
<tr>
<td>I/σI</td>
<td>17.6 (4.6)</td>
<td>26.6 (7.0)</td>
<td>17.9 (4.2)</td>
<td>9.6 (3.8)</td>
</tr>
<tr>
<td><strong>C. Phasing, SOLVE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30–2.8</td>
<td>25–2.8</td>
<td>25–2.8</td>
<td>25–3.1</td>
</tr>
<tr>
<td>Riso</td>
<td>14.0</td>
<td>15.2</td>
<td>15.2</td>
<td>17.0</td>
</tr>
<tr>
<td>Rano</td>
<td>4.5</td>
<td>6.5</td>
<td>6.5</td>
<td>10.1</td>
</tr>
<tr>
<td>Sites</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Occupancy (%)</td>
<td>19.9, 10.5</td>
<td>37.1</td>
<td>37.1</td>
<td>52.0</td>
</tr>
<tr>
<td>Peak height (σ)</td>
<td>38.2, 20.4</td>
<td>13.1</td>
<td>13.1</td>
<td>36.2</td>
</tr>
<tr>
<td>mean FOM (%)</td>
<td>39.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-score (σ)</td>
<td>16.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>D. Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P6₃</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30–1.90 (1.97–1.90)</td>
<td>25–2.80 (2.90–2.80)</td>
<td>25–3.10 (3.21–3.10)</td>
<td></td>
</tr>
<tr>
<td>R抗震</td>
<td>20.5 (22.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rfree (1792 reflections)</td>
<td>24.0 (28.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>3302</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water molecules</td>
<td>378</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rms deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond lengths (Å)</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond angles</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average B values (Å²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>16.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>23.13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - \langle I \rangle|}{\sum |I_{\text{obs}}|}, \quad R_{\text{iso}} = \frac{\sum |F_{\text{calc}} - F_{\text{obs}}|}{\sum F_{\text{calc}}}, \quad R_{\text{ano}} = \frac{\sum |F(+) - F(-)|}{\sum |F(+) + F(-)|}, \quad \text{Terwilliger & Berendzen, 1999).} \]
The three-dimensional structure of a superantigen-like protein, SET3, from a pathogenicity island of the Staphylococcus aureus genome

Vickery L. Arcus, Ries Langley, Thomas Proft, John D. Fraser and Edward N. Baker

J. Biol. Chem. published online June 24, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M203914200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2002/06/24/jbc.M203914200.citation.full.html#ref-list-1