

Virulence Factors from *Staphylococcus aureus*: Tools to Study Innate and Adaptive Immunity

Ries Langley¹, Vic Arcus² and John Fraser¹

¹Centre for Molecular Biodiscovery and Department of Molecular Medicine,
School of Medical Sciences, University of Auckland, New Zealand

²School of Biological Sciences, University of Auckland, New Zealand

Staphylococcus aureus is a Gram positive organism that lives within the body and on the skin as a persistent part of the commensal flora in 10 to 20% of healthy adults. At any one time up to 50% of the population can be colonised (1). Under certain conditions, *S. aureus* can evade the host regulatory mechanisms that keep it in its normally benign state, erupting into a florid infection that can rapidly develop into a serious and even life threatening condition. Diseases resulting from a staphylococcal infection include skin abscesses, bacteraemia, pneumonia, endocarditis, toxic shock syndrome, staphylococcal scarlet fever and scalded skin syndrome. *S. aureus* is also notorious for its ability to develop broad antibiotic resistance. The new strain resistant to vancomycin, VRSA, is resistant to all known antibiotics and emphasises that this organism poses a major threat to human health globally. Not surprisingly, *S. aureus* was a priority organism for genome sequencing and there are now four completed genomes available including three methicillin resistant strains (MRSA) NCTC8325, MW2 and Mu50, and the vancomycin resistant strain (VRSA) N315. The NCTC8325 genome sequence can be obtained from

Oklahoma's Advanced Center for Genome Technology <http://www.genome.ou.edu>, while those of MW2, Mu50 and N315 are all available from the GenBank ftp server <ftp://ncbi.nlm.gov/genomes>.

S. aureus is notorious not only for its ability to rapidly acquire antibiotic resistance, but also for the wide variety of factors it uses in its defence against host immunity. Some of these virulence factors are well known, while others have been uncovered following the sequencing of the genome. Not only are these molecules interesting in their own right, but they offer unique and exquisitely tuned tools to investigate the components and complexities of host innate and adaptive immunity. This review could not possibly cover all the known staphylococcal virulence factors. Instead, it provides an overview of the latest and most intriguing addition, a family of genes located within a pathogenicity island, a chunk of mobile DNA that contains a concentration of virulence factors. This new family was uncovered by comparative searching of the staphylococcal genomes using a conserved sequence of another well-known virulence factor, the superantigen (SAg). Surprisingly,

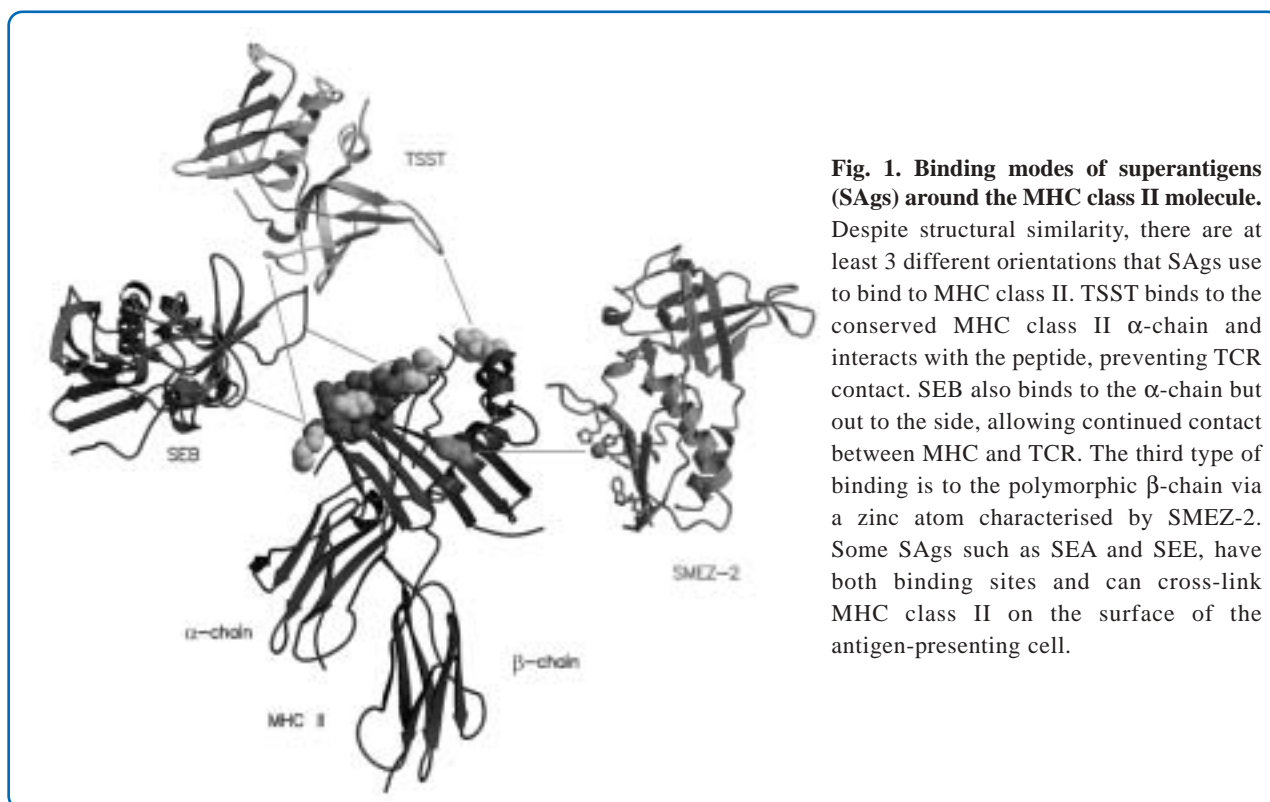


Fig. 1. Binding modes of superantigens (SAGs) around the MHC class II molecule. Despite structural similarity, there are at least 3 different orientations that SAGs use to bind to MHC class II. TSST binds to the conserved MHC class II α -chain and interacts with the peptide, preventing TCR contact. SEB also binds to the α -chain but out to the side, allowing continued contact between MHC and TCR. The third type of binding is to the polymorphic β -chain via a zinc atom characterised by SMEZ-2. Some SAGs such as SEA and SEE, have both binding sites and can cross-link MHC class II on the surface of the antigen-presenting cell.

despite significant structural similarity with SAGs, the Staphylococcal Enterotoxin-Like toxins (SETs) have quite different functions. This reflects how the organism has utilised common protein super-folds for a variety of roles based around virulence.

Bacterial superantigens

The best-studied staphylococcal virulence factors are the SAGs: a family of small secreted exotoxins that bind simultaneously to Major Histocompatibility Complex (MHC) class II molecules expressed on antigen-presenting cells (APC) of the adaptive immune system, and T-cell receptors expressed on T-lymphocytes, eliciting profound T-cell activation.

The term superantigen was originally coined by White *et al.* (2) to describe MHC-restricted antigens that are recognised by all $\alpha\beta$ T-cell receptors (TCR) bearing particular V β elements. SAGs are extremely potent, with some active at only a few thousand molecules per mL. During the process of normal antigen recognition, peptide antigen is presented in the groove of the two distal domains of MHC molecules to be recognised by specific TCR – a process that depends on the unique combination of the hypervariable domains of both TCR α -chain and β -chain. Thus, for any single MHC-peptide combination, a naive immune response relies on no more than 0.001% of the total T-cell repertoire. In contrast, SAGs cause a hyperimmune response. They bind to the sides of MHC class II molecules and TCR involving only the variable domain of the β -chain (3). SAG binding causes profound signalling via TCR, resulting in T-cell proliferation and cytokine release (4). Such binding also causes APC activation via MHC II signalling, to enhance inflammatory cytokine release (5). The massive Th1 cytokine and pro-inflammatory release can lead to hypotension, shock, organ failure, and death (6). As a consequence of SAG signalling, the T-cell is left in a state of anergy (7) where it can no longer respond to normal antigen, and can sometimes lead to death by apoptosis of the T-cell, if not rescued by second mediators such as IL-2 (8).

The relationships between the various SAGs are discussed below. Here we mention the variety of mechanisms employed by SAGs to bind to MHC class II. Some SAGs, such as SEB and TSST, bind to the invariant MHC class II α -chain while others, such as the streptococcal SAG SPE-C, bind to the other side of MHC class II to the polymorphic MHC class II β -chain via a zinc atom. A third binding mode exists for SAGs such as SEA and SEE. These combine both α -chain and β -chain MHC class II binding sites so that they bind to both sides of the MHC class II molecule and cause profound cross-linking of MHC class II on the surface of the APC (3) (**Fig. 1**).

Sequence alignment of the entire SAG family (**Fig. 2**) generates a tree that comprises a single outlier – toxic shock syndrome toxin-1 (TSST-1). This tree contains the more closely related staphylococcal enterotoxins (SE) A-E, and G-Q and the streptococcal pyrogenic exotoxins (Spe) A, C,

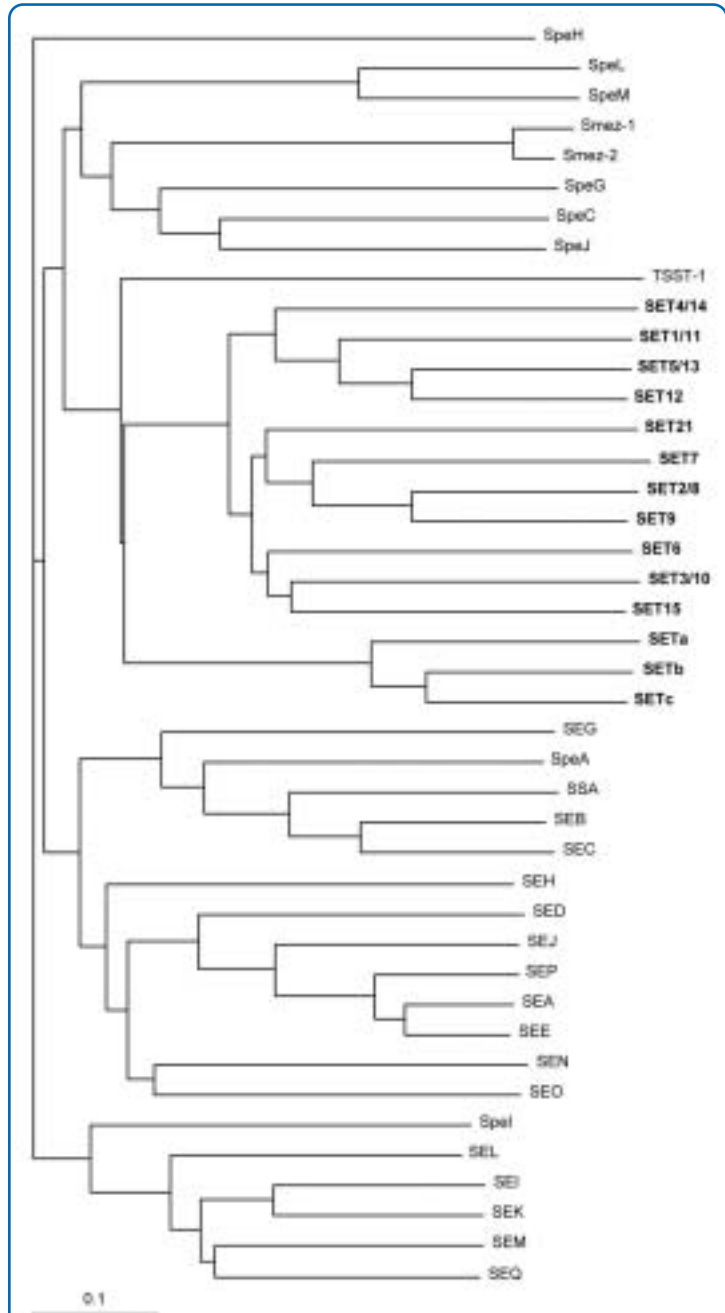


Fig. 2. Phylogenetic relationship between the SETs and the bacterial superantigens.

The phylogram shows that the SETs (bold font) form a subgroup with TSST-1. The tree was created from 1000 bootstraps of an amino acid sequence alignment generated using ClustalW.

G-M, SSA, and extremely potent streptococcal mitogenic exotoxin Z (SMEZ) which has many allelic forms (9). Overall amino acid sequence identity is low across the entire family (~20%) but the eleven 3D crystal structures that have now been determined reveal a basic conserved structure for all, including TSST-1. Each SAG is composed of a larger C-terminal domain with a β -grasp motif and a smaller N-terminal domain with an OB-fold motif (10). The latter is a common fold in bacterial proteins that bind oligomeric molecules such as DNA and polysaccharides (this is discussed in more detail below). A highly conserved alpha helix (α 4) within the C-terminal domain forms the central

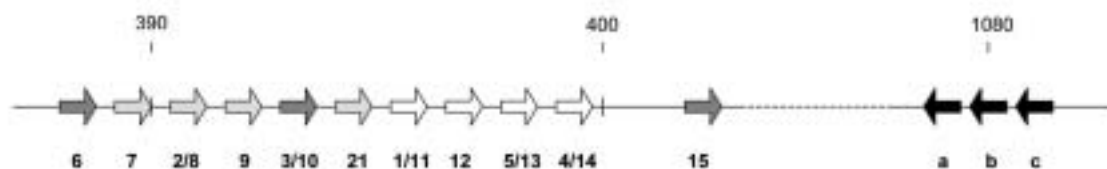


Fig. 3. Genomic localisation of the set genes.

Numbering above the diagram represents position within the genome of *S. aureus* NCTC8325 in kilobases. Numbering below the diagram represents *set* nomenclature. The shading of the individual *set* genes corresponds to the subgroup they belong to within the SET family (cf. Fig. 2).

core interface between the 2 domains. The PROSITE SAg motif K-X(2)-[LIVF]-X(4)-[LIVF]-D-X(3)-R-X(2)-L-X(5)-[LIV]-Y (PS00278) is located within this core α -helix.

The Staphylococcal Enterotoxin-Like (SET) genes

The *S. aureus* strain NCTC8325 was selected for sequencing at the University of Oklahoma's Advanced Center for Genome Technology (11). This genome was completed in 2001 and assembled into a contiguous circular sequence of 2,821,915 bp. Searching the NCTC8325 genome with the α 4 SAg sequence, uncovered 12 new SAg-like genes. They reside in two separate clusters within the genome and were named Staphylococcal Exotoxin-Like (SET) proteins (12). 26 members (SET1–SET26) have now been identified (12–14), although many of these proteins appear to be allelic variants of each other. For example SET1 (12), SET11 (13), and SET22 (14) are the same protein.

The SETs exhibit a high degree of identity in the α 4 region of SAg, but diverge significantly over the remainder of the molecule. Overall sequence identity between the SETs and the SAg is less than 20%, with greater conservation in the C-terminal half of the molecule. Sequence identity within the SET family ranges from 19–70%. Surprisingly, the SETs are most similar to TSST-1, and create a subgrouping on the alignment tree that includes TSST-1 (Fig. 2). The *set* cluster has since been identified as part of a pathogenicity island by Kuroda *et al.* (13) who completed the sequence of two new *S. aureus* isolates – Mu50 (meticillin resistant – MRSA) and N315 (vancomycin resistant strain – VRSA) in 2001 (13). Kuroda *et al.* identified three pathogenicity islands spread evenly across the genome. SaPI_n1 contains the TSST gene, SaPI_n2 contains the SETs (named the exotoxin island) (Fig. 3) as well as other potential virulence factors, and SaPI_n3 contains the majority of the SAg genes (named the enterotoxin island) (13).

Clusters of *set* genes have been identified in all the sequenced *S. aureus* genomes, but are absent in the published *Staphylococcus epidermidis* or the completed streptococcal genomes. Fitzgerald *et al.* (15) report the presence of the SET cluster in all 36 *S. aureus* strains they examined. This suggests a non-redundant function for each member. Furthermore, the presence of the SET genes on a pathogenicity island, their high degree of allelic variation, and their relatedness to the SAg, together with the fact that the SETs are secreted by *S. aureus* and there is abundant seroconversion against them (16) suggests that they target host defence proteins.

Structure and function of SETs and SAg

We have produced recombinant forms of all the SETs and performed structural and functional studies. Initial functional assays quickly determined that none of the SETs are SAg. None of the recombinant SETs bind MHC class II nor do they activate human T-cells *in vitro* (16). Despite the lack of SAg function, molecular modelling of SET1–SET4 against the structures of the SAg TSST-1 and SpeC suggested that they shared similar overall 3D structures (12). The experimental determination of the crystal structure of SET3 (16) has recently confirmed a high degree of similarity to TSST-1 (Fig. 4).

Protein super-folds in bacterial virulence factors – comparison between SAg and SETs

Structural comparison between SET3 and TSST-1 (Fig. 4) identifies 37 conserved residues that either contribute to the hydrophobic core, or with their side chains forming buried charge-charge interactions or hydrogen bonding with main chain atoms (16). These are likely to be required for structural stability. Surface residues are very different between the two molecules. Comparison of those SAg regions known to be MHC class II or TCR contact regions reveal no conservation whatsoever, consistent with our inability to identify any SAg-like function.

The N-terminal domain of the SAg and SETs is composed of 5 β -strands that form a concave β -barrel. In all SAg except TSST-1 there is a short α -helix at the end of the barrel. This domain resembles the oligosaccharide/oligonucleotide binding fold or OB (oligomer binding)-fold motif first described by Murzin (10). Other secreted bacterial toxins containing this motif include staphylococcal nuclease, and the B subunits of the AB₅ toxins, which include cholera toxin, heat-labile enterotoxins I and IIb, verotoxin, and pertussis toxin (10,17).

The OB-fold is used for DNA binding and carbohydrate binding respectively in these proteins and is thought to be a stable common fold able to accommodate sequence variation and alternative functions. Orengo *et al.* (18) suggest that the OB-fold represents a super-fold from which other folds have evolved. These are extra-stable folds that may have diverged from a common ancestor and retained similar topology despite changes in sequence and function. Alternatively these folds may have arisen by convergence of polypeptide chains to adopt more stable conformations. While no oligosaccharide or oligonucleotide binding has yet been observed for the SAg or SETs, the flexibility and core stability of this motif may allow for a wide variety of derivatives which culminate in their ability to bind MHC class II molecules.



Fig. 4. Similarity of 3D structures of TSST-1 and SET3.

Both TSST-1 (at right), a typical superantigen, and SET3 (at left), representative of the SETs, share the same basic fold, although the latter proteins do not have superantigen activity. Note that both have the core α -helix and a 2-domain structure comprising the larger C-terminal domain with its β -grasp fold (at left of each image) and the smaller N-terminal domain with its OB-fold structure (at right of each image).

The C-terminal domain, or domain A, is made up of a 4 stranded anti-parallel β -sheet backing on to the central α -helix (the β -grasp motif), plus 3 coplanar β -strands, and an additional α -helix. Part of domain A is the very N-terminal tail and α -helix of the molecule, which packs against the β -sheet. Domain A shares some structural features with the β -grasp motif present in the B domain of streptococcal protein G, and domain B of *Peptostreptococcus magnus* protein L. Both are the immunoglobulin binding domains or immunoglobulin binding proteins (17). Other proteins containing the β -grasp motif are ubiquitin and ubiquitin-like proteins (19), the ferredoxins (20), and the Ras binding domain of the protein kinase Raf1 (21). The β -grasp motif is another example of a stable super-fold (18).

Dependency of *S. aureus* on SET genes

Confirming that a gene codes for a virulence factor ultimately requires clinical evidence that a strain expressing the gene is more virulent in a natural infection than one without it. This is not always easily achieved, but several methods have been developed to isolate individual genes using knockout, gene tagging or entrapment methods then testing of the isolated gene mutant in an animal infection model. Schneider *et al.* (22) used green fluorescent protein (GFP) gene tagging of random chromosomal DNA fragments to identify *S. aureus* promoters. A library of GFP-active reporter constructs were expressed in *S. aureus* cells, which were subsequently subjected to a variety of environmental conditions likely to be encountered *in vivo* by a pathogen. Active promoters (detected by GFP production) were identified under different conditions and their natural downstream gene identified simply by searching the completed genome. Genes of interest were then rendered inactive by the introduction of an erythromycin resistance gene within the gene, and the mutant strain tested in an animal infection model.

This method has identified a number of genes, some of which are critical to the vegetative life of the organisms, or when deleted, severely limit the organisms growth rate. Of the latter type, *set15* was found to be a member. The *set15*

promoter was found to be active in conditions of divalent cation starvation and the *set15*⁻ mutant displayed a greater than 30-fold reduction in bacterial survival in a murine kidney abscess infection model (22). From this elegant study, at least 1 of the secreted SETs, SET15, has been shown to enhance the survival of *S. aureus*. It now remains to determine the biochemical function of SET15 and how it enhances bacterial survival. Already there are some tantalising clues to its function.

Possible functions of the SETs

Preliminary results using recombinant forms of the SETs to bind serum proteins have identified a range of molecular targets that are selectively bound by each SET. Some of these serum proteins have been identified as central components of the complement cascade while others appear to interfere with Fc receptor-mediated interactions by selectively binding to either the Fc region of immunoglobulins or the Fc receptors themselves. Further experiments are required to confirm this activity and the downstream effects of its action.

The sequencing of microbial genomes such as *S. aureus* has provided a wealth of new genes. Of the ~2500 genes identified within the single circular chromosome, over 40% have no identifiable function. One of the most intriguing features of the staphylococcal genome are the 3 pathogenicity islands that contain a cluster of genes whose products are almost certainly designed to target host defence molecules. Both the SAg genes and the SETs are located within these pathogenicity regions and have clearly evolved from the same ancestral gene. Whether the SETs preceded the SAg genes or vice versa remains to be determined, but they represent a clear example of how a common protein fold has been utilised and modified for different purposes. The challenge now is to identify the molecular targets of the SETs and their role in pathogenesis.

References

1. Lowy, F.D. (1998) *New Engl. J. Med.* **339**, 520-532
2. White, J., Herman, A., Pullen, A.M., Kubo, R., Kappler, J.W., and Marrack, P. (1989) *Cell* **56**, 27-35
3. Fraser, J., Arcus, V., Kong, P., Baker, E., and Proft, T. (2000) *Mol. Med. Today* **6**, 125-132
4. Miethke, T., Wahl, C., Regele, D., Gaus, H., Heeg, K., and Wagner, H. (1993) *Immunobiology* **189**, 270-284
5. Mehindate, K., Thibodeau, J., Dohlsten, M., Kalland, T., Sekaly, R.P., and Mourad, W. (1995) *J. Exp. Med.* **182**, 1573-1577
6. Kotb, M. (1995) *Clin. Microbiol. Rev.* **8**, 411-426
7. Miethke, T., Wahl, C., Heeg, K., and Wagner, H. (1993) *J. Immunol.* **150**, 3776-3784
8. Kawabe, Y., and Ochi, A. (1991) *Nature* **349**, 245-248
9. Proft, T., Moffatt, S.L., Weller, K.D., Paterson, A., Martin, D., and Fraser, J.D. (2000) *J. Exp. Med.* **191**, 1765-1776

References continued on page 18



Continued from page 14

A Toolbox of *S. aureus* Virulence Factors

10. Murzin, A.G. (1993) *EMBO J.* **12**, 861-867
11. Roe, B.A., Tian, H., Jia, S., Li, S., Lin, S., Kenton, S., Lai, H., White, J.D., Dorman, A., Najar, F.Z., Clifton, S., Worrell, V., and Iandolo, J.J. *Staphylococcus aureus* Genome Sequencing Project. University of Oklahoma. Funded by the NIH and the Merck Genome Research Institute
12. Williams, R.J., Ward, J.M., Henderson, B., Poole, S., O'Hara, B.P., Wilson, M., and Nair, S.P. (2000) *Infect. Immun.* **68**, 4407-4415
13. Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., *et al.* (2001) *Lancet* **357**, 1225-1240
14. Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., *et al.* (2002) *Lancet* **359**, 1819-1827
15. Fitzgerald, J.R., Sturdevant, D.E., Mackie, S.M., Gill, S.R., and Musser, J.M. (2001) *Proc. Nat. Acad. Sci. USA* **98**, 8821-8826
16. Arcus, V.L., Langley, R., Proft, T., Fraser, J.D., and Baker, E.N. (2002) *J. Biol. Chem.* **277**, 32274-32281
17. Mitchell, D.T., Levitt, D.G., Schlievert, P.M., and Ohlendorf, D.H. (2000) *J. Mol. Evol.* **51**, 520-531
18. Orengo, C.A., Jones, D.T., and Thornton, J.M. (1994) *Nature* **372**, 631-634
19. Buchberger, A., Howard, M.J., Proctor, M., and Bycroft, M. (2001) *J. Mol. Biol.* **307**, 17-24
20. Tsukihara, T., Fukuyama, K., Mizushima, M., Harioka, T., Kusunoki, M., Katsube, Y., Hase, T., and Matsubara, H. (1990) *J. Mol. Biol.* **216**, 399-410
21. Nassar, N., Horn, G., Herrmann, C., Scherer, A., McCor F., and Wittinghofer, A. (1995) *Nature* **375**, 554-560
22. Schneider, W.P., Ho, S.K., Christine, J., Yao, M., Marra, Hromockyj, A.E. (2002) *Infect. Immun.* **70**, 1326-1333

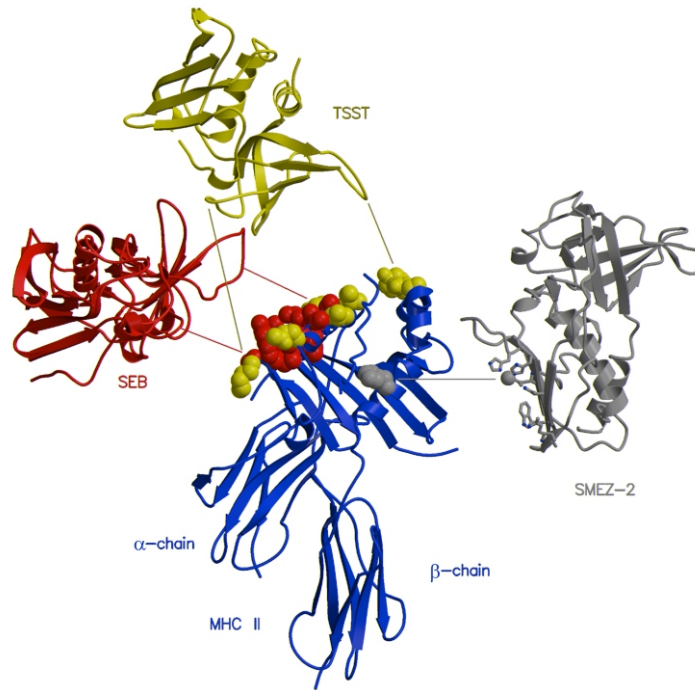


Fig. 1



Fig. 4

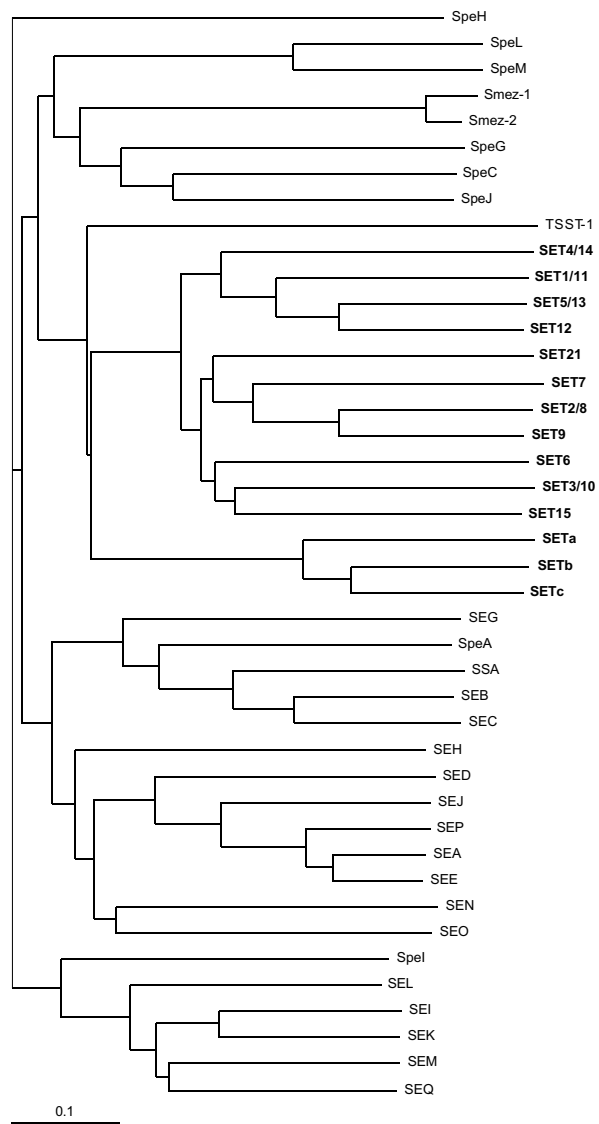


Fig. 2

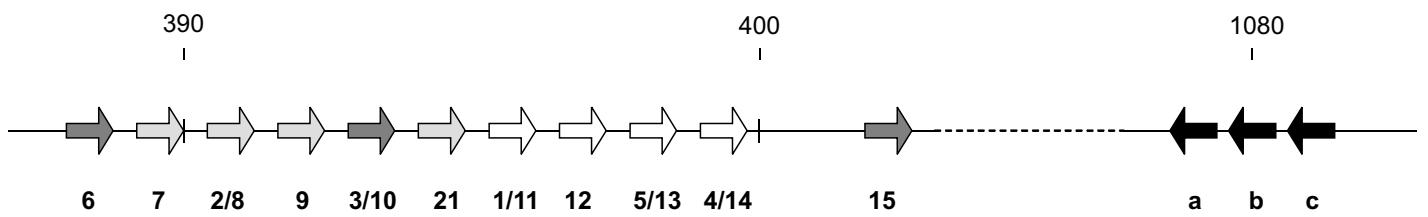


Fig. 3

SaPIIn2