The Crystal Structures of EAP Domains from *Staphylococcus aureus*

Reveal an Unexpected Homology to Bacterial Superantigens

Brian V. Geisbrecht‡§§, Brent Y. Hamaoka‡, Benjamin Perman‡, Adam Zemla‖ and Daniel J. Leahy¶§**

From the ‡Department of Biophysics and Biophysical Chemistry and the §Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205 and the ‖Lawrence Livermore National Laboratory, Livermore, California 94550

The Eap (extracellular adherence protein) of *Staphylococcus aureus* functions as a secreted virulence factor by mediating interactions between the bacterial cell surface and several extracellular host proteins. Eap proteins from different Staphylococcal strains consist of four to six tandem repeats of a structurally uncharacterized domain (EAP domain). We have determined the three-dimensional structures of three different EAP domains to 1.8, 2.2, and 1.35 Å resolution, respectively. These structures reveal a core fold that is comprised of an α-helix lying diagonally across a five-stranded, mixed β-sheet. Comparison of EAP domains with known structures reveals an unexpected homology with the C-terminal domain of bacterial superantigens. Examination of the structure of the superantigen SEC2 bound to the β-chain of a T-cell receptor suggests a possible ligand-binding site within the EAP domain (Fields, B. A., Malchiodi, E. L., Li, H., Ysern, X., Stauffacher, C. V., Schlievert, P. M., Karjalainen, K., and Mariuzza, R. (1996) *Nature* 384, 188–192). These results provide the first structural characterization of EAP domains, relate EAP domains to a large class of bacterial toxins, and will guide the design of future experiments to analyze EAP domain structure/function relationships.

*Staphylococcus aureus* is a widespread, persistent pathogen that causes a broad range of diseases in humans and animals, from simple wound infections to more severe conditions such as septicemia, endocarditis, and osteomyelitis (2). As is the case with many pathogens, initiation of sepsis can progress to severe conditions such as meningitis, peritonitis, and shock (3). The Eap system is an important component of the immune defenses of the host. It is responsible for the efficient entry and colonization of nonphagocytic host cells (13, 14). Thus, there is interest in understanding and evaluating Eap-mediated interactions as potential targets for novel antimicrobial therapies (5).

Although different receptors bind a variety of ligands in host tissues, a common feature is their ability to bind specifically to the large glycoproteins present in the extracellular matrix (e.g. fibronectin, collagen, and fibrinogen) (reviewed in Ref. 5). The 50–70-kDa Eap (extracellular adherence) protein of *S. aureus* was originally identified based on its ability to bind glomerular basement membranes (6), and more mechanistic studies have demonstrated that Eap is capable of binding a wide array of extracellular matrix proteins (7–9), including fibronectin, vitronectin, laminin, and the pro-inflammatory cell surface receptor intercellular adhesion molecule 1 (ICAM-1) (10). Despite the fact that the mature Eap protein lacks a transmembrane region, ~30% of the steady-state level of Eap was found associated with the bacterial cell surface (9). Further studies demonstrated that this association can occur through specific protein-protein interactions (11) and uncharacterized membrane structures (12) as well as through Eap multimerization (9). Because no other receptor has these properties, Eap is believed to serve a unique role in bacterial virulence by mediating interactions between abundant host ligands (e.g. fibronogen, fibronectin, and vitronectin) and the bacterial cell. These interactions result in bacterial agglutination within the host and are believed to contribute to *S. aureus* pathogenesis (9, 13, 14). Consistent with this idea, *S. aureus* strains that fail to produce Eap exhibit markedly lower affinity for eukaryotic cells (12–14) and lack the ability to efficiently enter and colonize nonphagocytic host cells (13, 14). Thus, there is interest in understanding and evaluating Eap-mediated interactions as potential targets for novel antimicrobial therapies (5).

Like many extracellular proteins, Eap exhibits a modular architecture and is believed to be comprised of tandem repeats of an ~110-residue domain (EAP domain) (15, 16). Depending on the strain examined, Eap is comprised of either four or six repeats of this domain, although extensive homology between these isoforms and nearly indistinguishable biochemical properties suggest that these proteins serve essentially the same physiological function (10, 17). These observations, combined with the wide variety of Eap ligands, imply that the individual EAP domains themselves are remarkably versatile in their ability to form protein-protein interactions. However, the biochemical properties of these domains are not well studied, nor do EAP domains share detectable sequence homology with any proteins of known structure.

As a first step toward a molecular understanding of EAP
domain-mediated protein-protein interactions, we determined the three-dimensional structures of three distinct EAP domain proteins to 1.8, 2.2, and 1.35 Å resolution, respectively. These structures consist of a 97-residue core fold composed primarily of a long α-helix laying diagonally across a five-stranded, mixed β-sheet. Comparison of EAP domains with other structures in the Protein Data Bank revealed an unexpected structural homology with the C-terminal domain of bacterial superantigens, a class of toxic proteins capable of massive stimulation of host lymphocytes by simultaneously binding both major histocompatibility class II proteins and T-cell receptors (TCRs)1 (18, 19). Examination of the crystal structure of the superantigen SEC bound to a TCR β-chain suggests a potential ligand-binding site within the EAP domain (1). Our results provide the first insights into the structure of EAP domains and identify a novel and unexpected relationship between two broad classes of bacterial virulence factors. These results provide a framework for future studies addressing the structure/function relationships of these and other EAP domain proteins.

**EXPERIMENTAL PROCEDURES**

**Molecular Cloning and Sequence Analysis**

*S. aureus* strain Mu50 was purchased from ATCC and subcultured in liquid LB medium using standard techniques. Genomic DNA was isolated from 1 liter of stationary phase bacteria according to the procedure originally described by Marmur (20), with the exception that lysostaphin (Sigma) was included in the lysis buffer according to the manufacturer’s suggestions. The crude preparation of genomic DNA was purified further by standard extraction and precipitation techniques and resuspended in double distilled H2O to a final concentration of 1.0 mg/ml (A260/λmax = 1.7).

The genes encoding EapH1 (GenBank™ accession number NP_372729) and EapH2 (GenBank™ accession number NP_371505) were identified in the Mu50 genomic sequence by BLAST searching with entire amino acid sequence of Eap (GenBank™ accession number NP_372462). The conceptual translation of each gene was analyzed for matching protocols (RESOLVE) (28, 29), an initial model of residues 31–141 of EapH1, residues 24–144 of EapH2, and residues 160–254 of Eap (the second EAP repeat, denoted EAP2) were amplified by PCR from Mu50 genomic DNA with SalI and NotI restriction endonuclease recognition sites at the 5’ and 3’ ends, respectively, subcloned into the corresponding sites of the pT7HMT Escherichia coli expression vector,2 and sequenced in their entirety. All of the sequence alignments were performed using the DNASTAR software suite.

**Protein Overexpression and Purification**

*E. coli* strain B834(DE3) was used for overexpression of all proteins. Cell growth for the production of native proteins was carried out at 37 °C as previously described (24). Following induction, the cells were harvested by centrifugation, resuspended in 50 ml of denaturing lysis buffer (0.1 M sodium P, (7.8), 6 M guanidinium HCl) per liter of original culture, lysed by stirring for 30 min at room temperature, and the solubilized, denatured proteins were separated from cellular debris by centrifugation (30 min at 30,000 × g). The clarified cell extract was applied by gravity flow to a 7.5-ml column of Ni2-nitrilotriacetic acid-Sepharose (Qiagen) at room temperature, and His-tagged proteins were isolated according to manufacturer’s suggestions. Following purification, recombinant proteins were refolded by rapid dilution2 and concentrated by native chelating chromatography according to the manufacturer’s suggestions (Qiagen).

The purified EAP domain fusion proteins were digested with recombinant TEV protease (Invitrogen) at 4 °C according to the manufacturer’s suggestions to avoid potential crystallization problems caused by protease digestion of the N-terminal fusion tag. When complete (as judged by SDS-PAGE), the samples were exchanged into 20 mM sodium formate (pH 3.5), applied to a 6-ml Resource S cation exchange column (Amersham Biosciences), washed free of loosely bound contaminants, and eluted with a gradient from 0 to 1 M NaCl over 10 column volumes. Fractions containing apparently homogenous EapH1, EapH2, or EAP2 were pooled, dialyzed once overnight against 4 liters of double distilled H2O, molecular weight cut off Slide-A-Lyzer cassette (Pierce Biotechnology), and lyophilized for storage.

For overexpression of selenomethionyl-derivatized mutant proteins, selective M9 minimal medium containing 50 mg/liter of seleno-l-methionine (Sigma) was prepared and used according to the general procedure of Beneken et al. (25). Each derivatized protein was purified from the cells in 1 liter of induced culture according to the same protocol used for native proteins, with the exception that all of the buffers were vacuum-degassed prior to use. The incorporation of selenomethionine in each mutant was complete as judged by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (PerkinElmer Life Sciences).

**Crystallization and Data Collection**

All of the crystals were grown in hanging drops at 20 °C by mixing 1 μl of protein solution (in water) with 1 μl of a 1:3 dilution of reservoir buffer in water and equilibrating over 750 μl of undiluted reservoir buffer. The crystals used for diffraction analysis were soaked briefly in a fresh aliquot of reservoir buffer and flash frozen in a gaseous nitrogen stream at −180 °C prior to data collection. All of the data were processed and scaled using the DENZO/SCALEPACK software suite (26). The data collection statistics are shown in Tables II and III. The crystallization conditions and related information for each protein is presented below.

EapH1—30 mg/ml protein was mixed with a reservoir buffer of 0.1 M sodium cacodylate (pH 6.7), 0.2 M zinc acetate, and 24% (w/v) polyethylene glycol 8000. Twinned, needle-like crystals appeared in ~3–4 days and continued to grow through 14 days. Diffraction quality crystals were grown by microseeding fragments of the needle crystals into identical drops that had been pre-equilibrated for 12 h. EapH1 crystals suitable for diffraction analysis grew in the space group P31 (unit cell dimensions: a = b = 38.55 Å, c = 69.23 Å). The diffraction data for MAD phasing were collected at Selenium edge, peak, and remote wavelengths for crystals of the I96M and L106M mutants using Beamline X4A of the National Synchrotron Light Source at Brookhaven National Laboratory. The diffraction data used for high resolution refinement were collected at Beamline X4A-C of the Advanced Photon Source at Argonne National Laboratory.

EapH2—30 mg/ml protein was mixed with a reservoir buffer of 0.1 M sodium cacodylate (pH 6.7), 0.2 M ammonium acetate, 30% (w/v) polyethylene glycol 8000, and 4% (w/v) 2,6-hexamethylenediamine. Twinned, needle-like crystals appeared within 2 days and achieved their full size in approximately 1 week. Microseeding produced a single crystal that grew in the space group P21 (unit cell dimensions: a = 50.01 Å, b = 61.97 Å, c = 67.98 Å; β = 105.21°) and contained two molecules in the asymmetric unit. The diffraction data were collected using CuKα radiation produced by a Rigaku RU-200 generator.

EAP2—13 mg/ml protein was mixed with a reservoir buffer of 0.1 M MES (pH 6.4), 0.1 M ammonium sulfate, and 24% (w/v) polyethylene glycol 8000. Single crystals appeared in 1 day, achieved their full size in approximately 1 week, and grew in the space group P321 (unit cell dimensions: a = b = 64.116 Å, c = 141.706 Å) with two molecules in the asymmetric unit. The diffraction data were collected at Beamline X4A of the National Synchrotron Light Source.

**Structure Solution and Refinement**

The two anticipated selenium sites in both the I96M and L106M mutants of EapH1 (one native and one introduced) were determined and the phases calculated using the program SOLVE (27). To generate the initial experimental phases, complete, three wavelength MAD data sets from both mutants were combined into a single, pseudo-multiple isomorphous replacement data set. The experimental values for the Figure of merit (m) are shown in Tables II and III. After the experimental phases were improved by solvent flattening and histogram matching protocols (RESOLVE) (28, 29), an initial model of residues 45–141 was built into 2.2 Å experimental electron density maps using the program O (30). The model was subjected to a single round of
simulated annealing and maximum-likelihood positional refinement using CNS and a native data set processed to 1.8 Å limiting resolution (31). Iterative rounds of model building and water addition were alternated with B-factor and positional refinement to yield the final model. This model consists of residues 43–141, 95 water molecules, and two ordered zinc ions. No electron density was observed for residues 31–42. The final model was analyzed by the program PROCHECK in CCP4 (32, 33). Greater than 90% of the residues were found in the most favorable regions of the Ramachandran plot, and no residues were found in the disallowed regions. The final refinement statistics are shown in Tables II and III.

The structure of EapH2 was solved by molecular replacement using the program MOLREP and the refined structure of EapH1 as a search model (32, 34). The top two molecular replacement solutions yielded models for both EapH2 molecules in the asymmetric unit, which were then subjected to a single round of rigid body refinement and simulated annealing prior to building a 1.35 Å limiting resolution initial model of residues 160–254 in both polypeptide chains using the program O (30). Noncrystallographic symmetry constraints were applied during the early stages of refinement and were removed once the $R_{\text{free}}$ value dropped below 40%. Several additional rounds of model rebuilding and water addition, alternated with positional and individual B-factor refinement, yielded the final model. This model consists of residues 160–254 for both protein molecules in the asymmetric unit, in addition to the residues Gly-Ser-Thr (an artifact of subcloning) at the N terminus, and 304 ordered solvent molecules. Superposition of all of the Ca positions from both polypeptide chains revealed no significant structural differences between the two molecules (r.m.s.d. = 0.29 Å). Again, the orientation of the noncrystallographic symmetry axis is unlikely to reflect any physiologically relevant dimerization because the EapH2 protein behaves as a monomer in solution at concentrations up to 2 mM as judged by analytical size exclusion chromatography (data not shown). The final model was analyzed by the program PROCHECK (32, 33). Greater than 90% of the residues were found in the most favorable regions of the Ramachandran plot, and no residues were found in the disallowed regions. The final refinement statistics are shown in Tables II and III.

The structure of EAP2 was solved by molecular replacement using the CCP4 program MOLREP and a single molecule from the refined structure of EapH2 as a search model (32, 34). The top molecular replacement solutions yielded models for both EAP2 molecules in the asymmetric unit, which were subjected to a single round of rigid body refinement and simulated annealing prior to building a 1.35 Å limiting resolution initial model of residues 160–254 in both polypeptide chains using the program O (30). Noncrystallographic symmetry constraints were applied during the early stages of refinement and were removed once the $R_{\text{free}}$ value dropped below 40%. Several additional rounds of model rebuilding and water addition, alternated with positional and individual B-factor refinement, yielded the final model. This model consists of residues 160–254 for both protein molecules in the asymmetric unit, in addition to the residues Gly-Ser-Thr (an artifact of subcloning) at the N terminus, and 304 ordered solvent molecules. Superposition of all of the Ca positions from both polypeptide chains revealed no significant structural differences between the two molecules (r.m.s.d. = 0.29 Å). Again, the orientation of the noncrystallographic symmetry axis is unlikely to reflect any physiologically relevant dimerization because the EapH2 protein behaves as a monomer in solution at concentrations up to 2 mM as judged by analytical size exclusion chromatography (data not shown). The final model was analyzed by the program PROCHECK (32, 33). Greater than 90% of the residues were found in the most favorable regions of the Ramachandran plot, and no residues were found in the disallowed regions. The final refinement statistics are shown in Tables II and III.

The structure of EAP2 was solved by molecular replacement using the CCP4 program MOLREP and a single molecule from the refined structure of EapH2 as a search model (32, 34). The top molecular replacement solutions yielded models for both EAP2 molecules in the asymmetric unit, which were subjected to a single round of rigid body refinement and simulated annealing prior to building a 1.35 Å limiting resolution initial model of residues 160–254 in both polypeptide chains using the program O (30). Noncrystallographic symmetry constraints were applied during the early stages of refinement and were removed once the $R_{\text{free}}$ value dropped below 40%. Several additional rounds of model rebuilding and water addition, alternated with positional and individual B-factor refinement, yielded the final model. This model consists of residues 160–254 for both protein molecules in the asymmetric unit, in addition to the residues Gly-Ser-Thr (an artifact of subcloning) at the N terminus, and 304 ordered solvent molecules. Superposition of all of the Ca positions from both polypeptide chains revealed no significant structural differences between the two molecules (r.m.s.d. = 0.29 Å). Again, the orientation of the noncrystallographic symmetry axis is unlikely to reflect any physiologically relevant dimerization because the EapH2 protein behaves as a monomer in solution at concentrations up to 2 mM as judged by analytical size exclusion chromatography (data not shown). The final model was analyzed by the program PROCHECK (32, 33). Greater than 90% of the residues were found in the most favorable regions of the Ramachandran plot, and no residues were found in the disallowed regions. The final refinement statistics are shown in Tables II and III.

The structure of EAP2 was solved by molecular replacement using the CCP4 program MOLREP and a single molecule from the refined structure of EapH2 as a search model (32, 34). The top molecular replacement solutions yielded models for both EAP2 molecules in the asymmetric unit, which were subjected to a single round of rigid body refinement and simulated annealing prior to building a 1.35 Å limiting resolution initial model of residues 160–254 in both polypeptide chains using the program O (30). Noncrystallographic symmetry constraints were applied during the early stages of refinement and were removed once the $R_{\text{free}}$ value dropped below 40%. Several additional rounds of model rebuilding and water addition, alternated with positional and individual B-factor refinement, yielded the final model. This model consists of residues 160–254 for both protein molecules in the asymmetric unit, in addition to the residues Gly-Ser-Thr (an artifact of subcloning) at the N terminus, and 304 ordered solvent molecules. Superposition of all of the Ca positions from both polypeptide chains revealed no significant structural differences between the two molecules (r.m.s.d. = 0.29 Å). Again, the orientation of the noncrystallographic symmetry axis is unlikely to reflect any physiologically relevant dimerization because the EapH2 protein behaves as a monomer in solution at concentrations up to 2 mM as judged by analytical size exclusion chromatography (data not shown). The final model was analyzed by the program PROCHECK (32, 33). Greater than 90% of the residues were found in the most favorable regions of the Ramachandran plot, and no residues were found in the disallowed regions. The final refinement statistics are shown in Tables II and III.

FIG. 1. The crystal structures of EapH1, EapH2, and EAP2. The crystal structures of EapH1 (A), EapH2 (B), and EAP2 (C) were solved and refined to 1.8, 2.2, and 1.35 Å limiting resolution, respectively. An alternative view of each structure achieved by rotating 90° counterclockwise in the viewing plane is shown on the right. Areas of distinct secondary structure were assigned by DSSP (21) and are colored green (β-helix), blue (α-helix), and purple (310-helix). The two ordered zinc ions in the EapH1 structure (A) are represented by orange spheres. The protein termini are labeled accordingly.
protein, as well as EAP2, were amplified from bacterial DNAs encoding the predicted mature forms of each
the second repeat in the Eap protein (EAP2), respectively
III). The EAP domain structures consist primarily of a four-
structure of EAP2 to 1.35 Å resolution (Fig. 1
resolution (Fig. 1
mine the structure of EapH2, which was refined to 2.2 Å
Purification, EapH1, EapH2, and EAP2 were crystallized by
Figure of merit
Heavy atom sites
2 2 3
No. of unique reflections
11,111 11,113 9,960
Completeness (%)
94.6 (64.2) 94.9 (65.8) 93.3 (58.3)
(⟨f(0)/σ(0)⟩)
16.3 (4.5) 15.8 (5.1) 17.3 (2.4)
Rmerge (%)\(^a\)
6.6 (19.4) 7.1 (16.4) 6.8 (28.3)

| ScMct EAPH1 (196M) |
|---------------------|
| Resolution limits (Å) | 30–1.8 | 30–1.8 | 30–1.8 |
| Wavelength (Å)        | 0.9792 | 0.9792 | 0.9687 |
| No. of reflections     | 39,859 | 40,286 | 37,925 |
| No. of unique reflections  | 10,111 | 10,113 | 9,960 |
| Completeness (%)       | 94.6 (64.2) | 94.9 (65.8) | 93.3 (58.3) |
| (⟨f(0)/σ(0)⟩)        | 16.3 (4.5) | 15.8 (5.1) | 17.3 (2.4) |
| Rmerge (%)\(^a\)      | 6.6 (19.4) | 7.1 (16.4) | 6.8 (28.3) |

| ScMct EAP106M |
|-----------------|
| Resolution limits (Å) | 30–2.0 | 30–2.0 | 30–2.0 |
| Wavelength (Å)        | 0.9794 | 0.9792 | 0.9725 |
| No. of reflections     | 28,815 | 31,465 | 32,203 |
| No. of unique reflections  | 7,395  | 7,468  | 7,531 |
| Completeness (%)       | 93.7 (58.8) | 94.8 (66.9) | 95.4 (70.1) |
| (⟨f(0)/σ(0)⟩)        | 16.0 (6.5) | 16.5 (6.9) | 17.3 (7.2) |
| Rmerge (%)\(^a\)      | 6.3 (12.2) | 6.4 (11.1) | 5.9 (11.1) |

| ScMct 1106M |
|-----------------|
| Complemd |

| | Edge | Peak | Remote |
|----------------|
| ScMcET EAPH1 (196M) |
| Resolution limits (Å) | 30–1.8 | 30–1.8 | 30–1.8 |
| Wavelength (Å)        | 0.9199 | 1.5418 | 0.9793 |
| No. of reflections     | 37,288 | 30,945 | 124,414 |
| No. of unique reflections  | 10,265 | 9,899  | 42,404 |
| Completeness (%)       | 99.1 (98.4) | 91.6 (51.3) | 88.8 (70.3) |
| (⟨f(0)/σ(0)⟩)        | 8.5 (2.6) | 10.2 (2.0) | 21.2 (1.9) |
| Rmerge (%)\(^a\)      | 12.5 (44.8) | 11.9 (31.7) | 4.4 (30.7) |
| Atoms modeled (protein/solvent) | 754/95 | 1,591/76 | 1,520/304 |
| r.m.s.d. for bonds (Å/angles (°)) | 0.006/1.3 | 0.010/1.5 | 0.020/2.0 |
| Average B for protein/solvent (Å\(^2\)) | 30.2/47.4 | 30.9/33.7 | 14.3/25.9 |
| B for zinc site 1/site 2 (Å\(^2\)) | 27.8/42.6 | 23.3/23.8 | 25.8/31.7 |

### RESULTS

The Three-dimensional Structures of Three EAP Domains—
Two uncharacterized EAP domain proteins were identified from *S. aureus* strain Mu50 by a BLAST search of the nonredundant protein data base for small proteins that contain EAP domains. These proteins, denoted EapH1 (GenBank\(^\text{TM}\) accession number NP_372729) and EapH2 (GenBank\(^\text{TM}\) accession number NP_37505), each contain a signal peptide at their N-terminus, followed by a unique stretch of 10–20 residues and then a single EAP domain. EapH1 and EapH2 share 47% identity with one another and are ~35 and 42% identical to the second repeat in the Eap protein (EAP2), respectively (Table I).DNAs encoding the predicted mature forms of each protein, as well as EAP2, were amplified from bacterial genomic DNA, subcloned, and expressed in *E. coli*. Following purification, EapH1, EapH2, and EAP2 were crystallized by hanging drop vapor diffusion methods (see “Experimental Procedures”).

The three-dimensional structure of EapH1 was determined using MAD data collected from crystals of selenomethionyl-substituted protein and refined to 1.8 Å limiting resolution (Fig. 1A and Tables II and III). The refined structure of EapH1 was used as a molecular replacement search model to determine the structure of EapH2, which was refined to 2.2 Å resolution (Fig. 1B and Tables II and III) and itself used as a molecular replacement search model to determine the structure of EAP2 to 1.35 Å resolution (Fig. 1C and Tables II and III). The EAP domain structures consist primarily of a four-turn α-helix laying diagonally across one face of a five-stranded, mixed β-sheet and resemble an open hand holding a roll of coins (Fig. 1, rotated).

Whereas the structures of all three EAP domains superimpose well (Table IV), two regions on either edge of the β-sheet exhibit notable structural divergence between these proteins (Fig. 2). The first of these regions lies between the β1 and β2 strands. In the case of EapH1 and EapH2, a short, additional β strand (denoted β1’ in Fig. 1, A and B) is found; this element is absent from EAP2, although strand β2 is extended by two residues toward the N-terminus in this protein. The second and more significant region of structural divergence is found between strands β5 and β6. In case of EapH1 and EapH2, the β5–β8 region consists of an ordered loop that contains a single turn of 3\(_{10}\)-helix (Fig. 1). This element is positioned roughly in the middle of the loop in EapH1 (Fig. 1A) but immediately follows strand β5 in EapH2 (Fig. 1B). This type of helical turn is not present in the structure of EAP2 (Fig. 1C). In addition to adopting distinct structures, the residues comprising these regions generally exhibit more flexibility as evidenced by above average temperature factors when compared with the remainder of each molecule. In the case of β1β2, these values are 39.9, 39.6, and 30.2 Å\(^2\), whereas for β5β6 they are 40.7, 31.1, and 20.6 Å\(^2\), where the mean temperature factors for the protein atoms are 30.2, 30.9, and 14.3 Å\(^2\) for the EapH1, EapH2, and EAP2 structures, respectively.

EAP Domains Share Unexpected Structural Homology with Bacterial Superantigens—The EAP domain structures were compared with structures in the Protein Data Bank by using...
The Structure of S. aureus EAP Domains

Fig. 2. Least squares superposition of EapH1, EapH2, and EAP2 reveals regions of structural divergence. The Ca backbone traces of EapH1 (blue), EapH2 (green), and EAP2 (purple) were superimposed by least squares fitting using the program O (30) and presented at both 0° (A) and 90° (B) views as in Fig. 1. The main regions of divergence between these structures occur in the loops between strands β1β2 and β5β6.

Discussion

To further our understanding of EAP domain structure and function, we have determined the crystal structures of three EAP domains, including the second repeat from S. aureus Eap. The results presented here represent the first structural characterization of EAP repeat proteins and show that these domains adopt a β-grasp fold composed of 97 residues. This fold is shared by the C-terminal domains of bacterial superantigens and establishes the unexpected result that superantigens and EAP domain proteins are members of a diverse superfamily of secreted bacterial toxins and virulence factors.

An attractive explanation for this structural and functional relationship is that EAP domains are a modular scaffold ideally suited to the formation of a wide variety of protein-protein interactions. Consistent with this hypothesis, the structural classification of proteins (SCOP) database indicates that the ability to participate in protein-protein interactions is a general feature of β-grasp proteins (40) (scop.bic.nus.edu.sg/data/scop.b.e.bi.html).

Functionally characterized Eap proteins are composed of multiple copies of the EAP domain, although it is not known whether these domains interact with one another or behave as physically discrete modules. Primary sequence analysis, biochemical data, and the structures presented here support the latter model, however. First, the individual domains in Eap are separated by positively charged, 10–14-amino acid stretches that are predicted to be random coil by secondary structure prediction algorithms (e.g. Ph.D. (41) or PSI-PRED (42)). Second, the full-length Eap protein is sensitive to protease treatment and is readily digested into soluble, protease-stable fragments corresponding to individual EAP domains.3 This suggests that these linker residues are solvent-exposed and/or accessible. Finally, EAP domains assume a fold where the N- and C-terminal residues are located at the opposite sides of the molecule. This topology is ideally suited to linking modular domains in succession and facilitates the formation of extended, rodlike structures. Similar arrangements are seen for fibronectin type III repeats and immunoglobulin domains, which are commonly found in extended arrays (43). This topology, along with the points mentioned above, suggests that individual EAP domains are unlikely to interact with one another in an intramolecular fashion and that each domain acts as a distinct binding module to create polyvalent EAP proteins capable of interacting with a wide variety of ligands and surfaces simultaneously. This secreted, molecular “Velcro” could serve to concentrate regions of bacterial growth and ultimately promote the infection of host cell layers. In this respect, one lingering question is whether the individual domains in Eap are responsible for binding to specific ligands or, alternatively, if individual domains are capable of binding a diverse set of ligands.

The homology between EAP domains and the C-terminal domain of bacterial superantigens suggests that the β5β6-interstrand region may be an important determinant of the specificity of EAP domain-mediated protein-protein interactions, and there are several lines of evidence that support this hypothesis. First, the crystal structure of superantigen SEC2 bound to a TCR β-chain revealed that residues lying within the

3 B. V. Geisbrecht and D. J. Leahy, unpublished observations.
the β5β6-interstrand region are critical in forming this complex (1).

Second, residues affecting the TCR β-chain binding specificity of superantigen SEA are found in the N-terminal portion of this region (44). Finally, examination of mutagenesis data in light of the TSST-1 crystal structure has implicated residues within the β5β6-interstrand region in toxin function (45). During the course of the studies presented here, we observed that the β5β6-interstrand region is the main area of sequence and structural divergence within this family of secreted virulence factors and toxins. This suggests that the β5β6-interstrand region is able accommodate a variety of amino acid sequences and structures to form a potentially broad repertoire of ligand-binding sites. The high resolution structures of the EAP domains reported here serve as a foundation for future experiments to address this question directly.

During the original characterization of S. aureus Eap, Jonsson et al. (15) reported that the individual EAP subdomains share a short region of homology with the N-terminal region of the β-subunit of the major histocompatibility class II proteins. Specifically, residues 3–33 in the mature major histocompatibility class II protein β chain were found to be homologous to residues corresponding to positions 93–123, 101–131, and 206–236 in the mature forms of EapH1, EapH2, and Eap, respectively. However, this region of homology is not detectable by
main proteins, such as EapH1, cannot be ruled out at this time.

Acknowledgments—We are grateful to the High-Throughput Crystallographic Laboratory of the Hauptmann-Woodward Institute (Buffalo, NY) for providing the initial crystals of EAP2. We also thank K. Ramay, S. Boyarin, P. Roescher, and J. McLeLLan for assistance with x-ray data collection, R. Abramowizit and X. Yang of Beamline X4-A at the National Synchrotron Light Source for technical assistance during MAD data collection, and M. Bebl of Bio-CARS Beamline 14BM-C at the Advanced Photon Source for additional native data collection.

REFERENCES

1. Fields, B. A., Malchiodi, E. L., Li, H., Ysern, X., Staffier, C. V., Schievert, P. M., Karjulainen, K., and Marruzzo, R. (1998) Nature 394, 188–192
2. Lowe, F. D. (1998) N. Engl. J. Med. 338, 503–502
3. Kronvall, G., and Jonsson, K. (1999) J. Mol. Recognit. 12, 1–7
4. Patti, J. M., Allen, B. L., McGavin, M. J., and Hook, M. (1999) Annu. Rev. Microbiol. 43, 585–617
5. Flick, J. I. (1999) Mol. Med. Today 5, 532–537
6. Youssif, Y., Mertz, A., Batsford, S. R., and Vogt, A. (1991) Zentbl. Bakteriol. 214, (suppl.) 168–169
7. McGavin, M. H., Krajewska-Pietrasik, D., Ryden, C., and Hook, M. (1993) Infect. Immun. 61, 2479–2485
8. Boden, M. K., and Flick, J. I. (1992) Microb. Pathog. 12, 289–298
9. Palma, M., Haggag, A., and Flick, J. I. (1999) J. Bacteriol. 181, 2840–2845
10. Chavakis, T., Hussain, M., Kanase, S. M., Peters, G., Brezel, R. G., Flock, J.-I., Herrmann, M., and Preissner, K. T. (2002) Nat. Med. 8, 687–693
11. Koc, M., and Flock, J.-I. (2001) J. Bacteriol. 183, 3999–4003
12. Kreikemeyer, B., McDevitt, D., and Podbioloski, A. (2002) Int. J. Med. Microbiol. 292, 283–295
13. Hussain, M., Haggag, A., Heimann, C., Peters, G. G., and Flock, J.-I., Herrmann, M., and Preissner, K. T. (2004) Infect. Immun. 72, 532–537
14. Haggag, A., Hussain, M., Lonnes, H., Herrmann, M., Nornby-Tegladu, A., and Flock, J.-I. (2005) Infect. Immun. 73, 2310–2317
15. Janossy, K., McDevitt, D., Harraghy, N., Peters, G. G., and Hook, M. (1995) J. Biol. Chem. 270, 21457–21460
16. Harraghy, N., Haggag, A., Chavakis, T., Sinha, B., Herrmann, M., and Flock, J.-I. (2005) Microbiology 149, 2701–2707
17. Hussain, M., Becker, K., von Eff C., Peters, G., and Herrmann, M. (2001) Clin. Diag. Lab. Immunol. 8, 1271–1276
18. Marrack, P., and Kappler, J. (1990) Science 246, 685–686
19. Herrmann, A., Kappler, J. W., Marrack, P., and Pullen, A. M. (1991) Ann. Rev. Immunol. 9, 745–772
20. Marmar, J. (1961) J. Biol. Chem. 236, 298–218
21. Rabson, W., and Sander, C. (1983) Biopolymers 22, 2577–2637
22. Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P., and Hendrickson, W. A. (1994) Proteins 18, 49–54
23. Geisbrecht, B. V., Dowd, K. A., Barfield, R. W., Longo, P. A., and Leahy, D. J. (2005) J. Biol. Chem. 280, 32561–32568
24. Geisbrecht, B. V., Zhang, D., Schulz, H., and Gould, S. J. (1999) J. Biol. Chem. 274, 21797–21803
25. Brozek, J., Tu, J. C., Xiao, B., Nuriya, M., Yuan, J. P., Wolfr, P. F., and Leahy, D. J. (2000) Nature 400, 143–154
26. Otwinowski, Z. (1993) in Proceedings of the CCP4 Study Weekend: Data Collection and Processing (Sawyer, L., Isacso, N., and Bailey, E., ed) pp. 56–62, Daresbury Laboratory, Hertfording, UK
27. Terwilliger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 849–861
28. Terwilliger, T. C. (2000) Acta Crystallogr. Sect. D Biol. Crystallogr. 56, 965–972
29. Terwilliger, T. C. (2002) Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 44–44
30. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. Sect. D Biol. Crystallogr. 47, 110–119
31. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Gross-Kusnstev, R. W., Jiang, J.-S., Kuszewski, J., Nigles, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonsson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 906–921
32. Collaborative Computational Crystallography Project 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 766–767
33. Laskowsk, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
34. Vanin, A., and Teplyakov, A. (1997) J. Appl. Crystallogr. 30, 1022–1205
35. Holm, L., and Sander, C. (1993) J. Mol. Biol. 233, 123–138
36. Keaert, C. A., Mitchell, D. T., Murray, D. L., Pinheiro, D. M., and Hook, M. (1996) J. Mol. Biol. 261, 7194–7202
37. Papageorgiou, A. C., Acharya, K. R. (1998) J. Mol. Biol. 287, 61–79
38. Roussel, A., Anderson, B. F., Baker, H. M., Frider, J. D., and Baker, E. N. (1997) Nat. Struct. Biol. 4, 635–643
39. Papageorgiou, A. C., and Acharya, K. R. (1997) Structure 5, 991–996
40. Murzin, A. G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995) J. Mol. Biol. 247, 536–540
41. Rost, B., and Sander, C. (1993) J. Mol. Biol. 222, 584–599
42. Jones, D. T. (1997) J. Mol. Biol. 292, 195–202
43. Lebl, D. J., Aukhil, I., and Erickson, H. P. (1998) Cell 84, 155–164
44. Irwin, M. J., Hadd, R. H., and Gascoigne, N. J. R. (1992) Nature 359, 841–843
45. Acharya, K. R., Pasadhesha, E. F., Jones, E. Y., Harlow, K., Stuart, D. I., Brehm, R. D., and Tranter, H. S. (1994) Nature 367, 94–97

FIG. 4. Structural analysis of superantigen complexes implicates the βββ-interstrand region as a main determinant of specificity in EAP domain-mediated protein–protein interactions. The structure of superantigen SEC2 complexed with the β-chain of a mammalian T-cell receptor (TCR-β) is shown (1). The TCR molecule is drawn in lavender, whereas the EAP-like C-terminal domain is colored according to secondary structure elements. The N-terminal domain of SEC is omitted for clarity.
46. Petersson, K., Hakansson, M., Nilsson, H., Forsberg, G., Svensson, L. A., Liljas, A., and Walse, B. (2001) EMBO J. 20, 3306–3312
47. Li, Y., Li, H., Dimasi, N., McCormick, J. K., Martin, R., Schuck, P., Schievert, P. M., and Mariuzza, R. A. (2001) Immunity 14, 93–104
48. Papageorgiou, A. C., Acharya, K. R., Shapiro, R., Passalacqua, E. F., Brehm, R. D., and Tranter, H. S. (1995) Structure 3, 769–779
49. Al-Daccak, R., Mehindate, K., Damdoumi, F., Etongue-Mayer, P., Nilsson, H., Antonsson, P., Sundstrom, M., Dohlsten, M., Sekaly, R. P., and Mourad, W. (1998) J. Immunol. 160, 225–232
50. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
51. Merritt, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524
52. Zemla, A. (2003) Nucleic Acids Res. 31, 3370–3374