

# The Crystal Structures of EAP Domains from *Staphylococcus aureus* Reveal an Unexpected Homology to Bacterial Superantigens\*

Received for publication, November 1, 2004, and in revised form, January 31, 2005  
Published, JBC Papers in Press, February 3, 2005, DOI 10.1074/jbc.M412311200

Brian V. Geisbrecht<sup>‡§¶</sup>, Brent Y. Hamaoka<sup>‡</sup>, Benjamin Perman<sup>‡</sup>, Adam Zemla<sup>||</sup>,  
and Daniel J. Leahy<sup>‡§\*\*</sup>

From the <sup>‡</sup>Department of Biophysics and Biophysical Chemistry and the <sup>§</sup>Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205 and the <sup>||</sup>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  $\alpha$ -helix lying diagonally across a five-stranded, mixed  $\beta$ -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  $\beta$ -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 *S. aureus* infection requires colonization of unique microenvironments within the host. To facilitate host colonization, *S. aureus* expresses a diverse array of virulence-determining factors, including exoenzymes, toxins, and numerous protein adhesins. These adhesins (termed Receptins (3, 4) or MSCRAMMs (4) (for microbial surface compo-

nents recognizing adhesive matrix molecules)) are believed to contribute to the initiation and propagation of infections by precluding bacterial clearance by physical forces (e.g. blood flow, coughing, etc.) and providing protection against host immune responses (4).

Although different receptins 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 receptin has these properties, Eap is believed to serve a unique role in bacterial virulence by mediating interactions between abundant host ligands (e.g. fibrinogen, 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

\* This work was funded by the Howard Hughes Medical Institute. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (codes 1YN3, 1YN4, and 1YN5; Refs. 50 and 51) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org>).

¶ Postdoctoral fellow of the Howard Hughes Medical Institute. Present address: Division of Cell Biology and Biophysics, University of Missouri-Kansas City, 5100 Rockhill Rd., Kansas City, MO 64110.

\*\* Assistant Investigator of the Howard Hughes Medical Institute. To whom correspondence should be addressed: Dept. of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe St., Baltimore, MD 21205. Tel.: 410-614-2533; Fax: 410-614-8839; E-mail: [dleahy@jhmi.edu](mailto:dleahy@jhmi.edu).

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  $\alpha$ -helix laying diagonally across a five-stranded, mixed  $\beta$ -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)<sup>1</sup> (18, 19). Examination of the crystal structure of the superantigen SEC bound to a TCR  $\beta$ -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 H<sub>2</sub>O to a final concentration of 1.0 mg/ml ( $A_{260}/A_{280} = 1.7$ ).

The genes encoding EapH1 (GenBank<sup>TM</sup> accession number NP\_372729) and EapH2 (GenBank<sup>TM</sup> accession number NP\_371505) were identified in the Mu50 genomic sequence by BLAST searching with entire amino acid sequence of Eap (GenBank<sup>TM</sup> accession number NP\_372462). The conceptual translation of each gene was analyzed for the highest probability signal peptide cleavage site using the SignalP algorithm Gram-positive option ([www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/)). DNAs encoding 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,<sup>2</sup> and sequenced in their entirety.

Site-directed mutagenesis was used to insert additional methionine residues in the EapH1 protein to facilitate phasing by multiwavelength anomalous dispersion (MAD) (22). The codons for Val<sup>91</sup>, Ile<sup>96</sup>, Leu<sup>106</sup>, and Val<sup>116</sup> were mutated independently to encode methionine by a two-step, megaprimer method to create four independent proteins with single-site methionine substitutions (23). Mutant DNAs encoding each variant protein were subcloned into pT7HMT 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<sub>i</sub> (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 Ni<sup>2+</sup>-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 dilution<sup>2</sup> 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 the presence of the N-terminal fusion tag. When digestion was 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 H<sub>2</sub>O using a 3500 molecular weight cut-off Slidealyzer 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  $\mu$ l of protein solution (in water) with 1  $\mu$ l of a 1:3 dilution of reservoir buffer in water and equilibrating over 750  $\mu$ l 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 P3<sub>2</sub> (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 14BM-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-hexanediol. 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 P2<sub>1</sub> (unit cell dimensions:  $a = 50.01$  Å,  $b = 31.97$  Å,  $c = 67.98$  Å;  $\beta = 105.21^\circ$ ) and contained two molecules in the asymmetric unit. The diffraction data were collected using CuK $\alpha$  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 within 1 day, achieved their full size in approximately 1 week, and grew in the space group R3 (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 final 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 ( $\langle m \rangle$ ) 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

<sup>1</sup> The abbreviations used are: TCR, T-cell receptor; MAD, multiwavelength anomalous diffraction; MES, 4-morpholineethanesulfonic acid; r.m.s.d., root mean square deviation.

<sup>2</sup> B. V. Geisbrecht, manuscript in preparation.

TABLE I  
Sequence similarities between the novel proteins EapH1 and EapH2 and the four subdomains of the Eap protein from *Staphylococcus aureus* Mu50

Sequences were aligned using the DNASTAR software suite and the percentage of sequence identity (top) and similarity (bottom) between the domains is presented. Similarity is defined as the percentage of identity plus the percentage of generally accepted conservative substitutions (*e.g.* Lys to Arg, Glu to Asp, etc.). The dashes indicate where values are not applicable.

	EapH1	EapH2	EAP1	EAP2	EAP3	EAP4
EapH1	—	47.0	30.0	35.0	25.0	33.0
EapH2	47.5	—	38.3	43.6	31.0	41.7
EAP1	31.6	42.1	—	83.1	48.4	56.8
EAP2	36.8	47.4	83.2	—	51.6	62.1
EAP3	26.3	33.7	48.4	51.6	—	58.9
EAP4	34.7	45.3	56.8	62.1	58.9	—

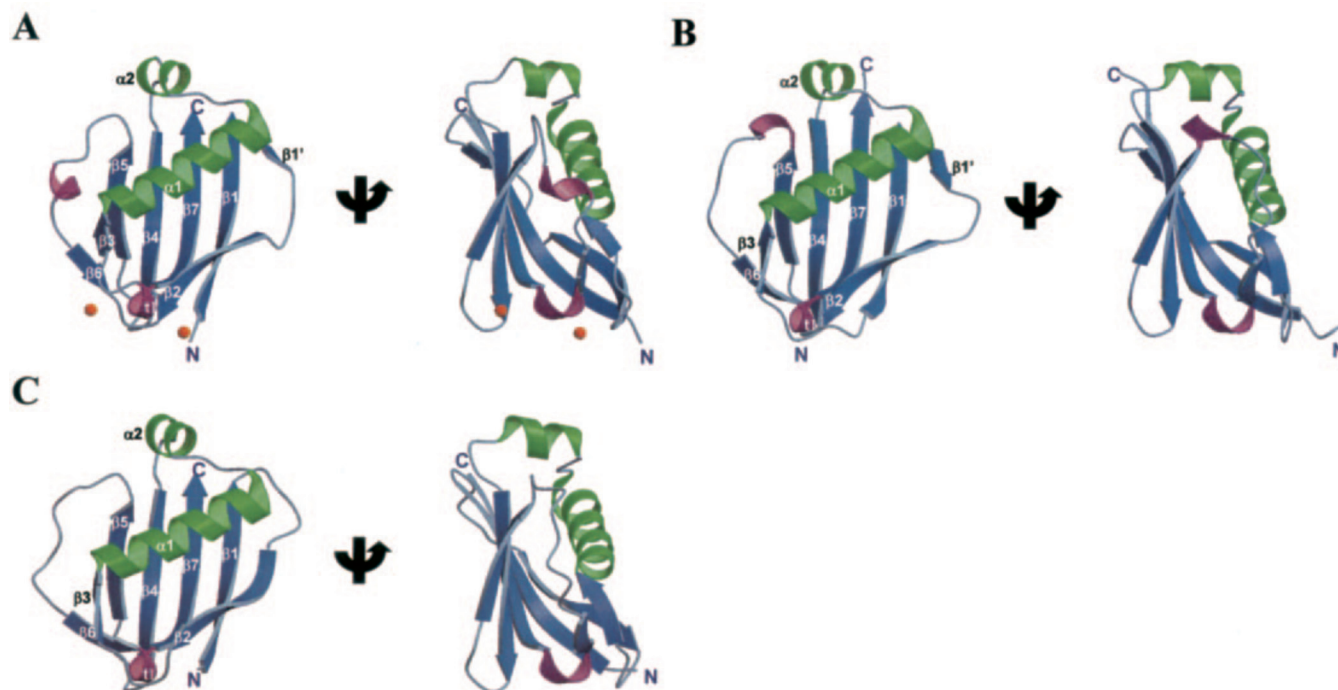


FIG. 1. The crystal structures of EapH1, EapH2, and EAP2. The crystal structures of EapH1 (A), EapH2 (B), and EAP2 (C) were solved and refined to 1.8, 2.2, and 1.35 Å limiting resolution, respectively. An alternative view of each structure achieved by rotating 90° counterclockwise in the viewing plane is shown on the right. Areas of distinct secondary structure were assigned by DSSP (21) and are colored green ( $\alpha$ -helix), blue ( $\beta$ -sheet), and purple ( $3_{10}$ -helix). The two ordered zinc ions in the EapH1 structure (A) are represented by orange spheres. The protein termini are labeled accordingly.

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 simulated annealing prior to building a 2.2 Å limiting resolution initial model of residues 45–144 in both polypeptide chains using the program O (30). Several rounds of model rebuilding and water addition, alternated with positional and individual B-factor refinement, yielded the final model. This model consists of residues 42–144 for the first protein molecule in the asymmetric unit, residues 45–144 for the second protein molecule, and 76 ordered solvent molecules. Superposition of all of the  $C\alpha$  positions of these polypeptide chains revealed no significant structural differences between the two molecules (r.m.s.d. = 0.43 Å). The orientation of the noncrystallographic symmetry axis is not likely to reflect any physiologically rele-

vant 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  $C\alpha$  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 dimer-



TABLE II

Data collection for experimental MAD phasing and structure solution  
The values in parentheses are for the highest resolution shell.

SeMet EAPH1 (L106M)			
	Edge	Peak	Remote
Resolution limits (Å)	30–1.8	30–1.8	30–1.8
Wavelength (Å)	0.9792	0.9793	0.9687
No. of reflections	39,859	40,286	37,525
No. of unique reflections	10,111	10,113	9,960
Completeness (%)	94.6 (64.2)	94.9 (65.8)	93.3 (58.3)
$\langle I \rangle / \langle \sigma I \rangle$	16.3 (4.5)	15.8 (5.1)	17.3 (2.4)
$R_{\text{merge}} (\%)^a$	6.6 (19.4)	7.1 (16.4)	6.6 (28.3)
SeMet EAPH1 (I96M)			
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	7,466	7,531
Completeness (%)	93.7 (58.8)	94.8 (66.9)	95.4 (70.1)
$\langle I \rangle / \langle \sigma I \rangle$	16.0 (6.5)	16.5 (6.9)	17.3 (7.2)
$R_{\text{merge}} (\%)^a$	6.3 (12.2)	6.4 (11.1)	5.9 (11.1)
SeMet I106M SeMet I96M Combined			
Heavy atom sites	2	2	3
Figure of merit $\langle m \rangle$	0.41	0.43	0.53

<sup>a</sup>  $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$ , where  $I_i(h)$  is the  $i$ th measurement of reflection  $h$ , and  $\langle I(h) \rangle$  is a weighted mean of all measurements of  $h$ .

ization because the EAP2 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 95% 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.

## 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<sup>TM</sup> accession number NP\_372729) and EapH2 (GenBank<sup>TM</sup> 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  $\alpha$ -helix laying diagonally across one face of a five-stranded, mixed  $\beta$ -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  $\beta$ -sheet

TABLE III

Data collection for model building and refinement

The values in parentheses are for the highest resolution shell.

	EapH1	EapH2	EAP2
Resolution limits (Å)	30–1.8	30–2.2	30–1.35
Wavelength (Å)	0.9199	1.5418	0.9793
No. of reflections	37,288	30,945	124,414
No. of unique reflections	10,268	9,899	42,404
Completeness (%)	99.1 (98.4)	91.6 (51.3)	88.8 (70.3)
$\langle I \rangle / \langle \sigma I \rangle$	8.5 (2.6)	10.2 (2.0)	21.2 (1.9)
$R_{\text{merge}} (\%)^a$	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 <sup>2</sup> )	30.2/47.4	30.9/33.7	14.3/25.9
B for zinc site 1/site 2 (Å <sup>2</sup> )	27.8/42.3		
$R_{\text{cryst}}/R_{\text{free}} (\%)^b$	23.3/23.8	25.8/31.7	21.8/23.0

<sup>a</sup>  $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$ , where  $I_i(h)$  is the  $i$ th measurement of reflection  $h$ , and  $\langle I(h) \rangle$  is a weighted mean of all measurements of  $h$ .

<sup>b</sup>  $R = \sum_h |F_{\text{obs}}(h) - F_{\text{calc}}(h)| / \sum_h |F_{\text{obs}}(h)|$ .  $R_{\text{cryst}}$  and  $R_{\text{free}}$  were calculated from the working and test reflection sets, respectively. The test set constituted 5% of the total reflections not used in refinement.

TABLE IV

Local-global alignment parameters for EapH1, EapH2, EAP2, and the superantigen TSST-1

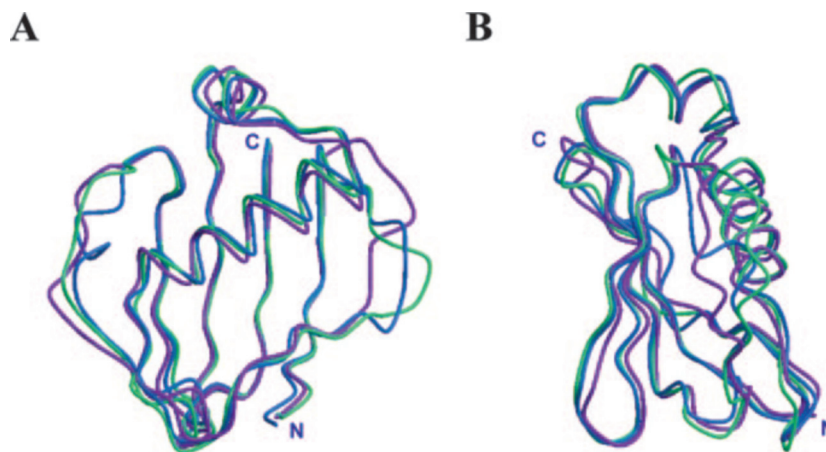
The structures were subjected to local-global alignment (52) to quantitatively assess the  $\alpha$  superpositions of the EAP domain(s) and TSST-1 structures. NS, number of residues in the refined structure; NT, number of residues in the reference structure; N, number of residues that superimpose within 2.5 Å; r.m.s.d.-N, root mean square deviation of N residues that fall within 2.5 Å.

	NS	NT	N ( $\leq 2.5$ Å)	r.m.s.d.-N
				Å
Local-global alignment using EAP2 as a reference				
EAP2	98	98	98	0
EapH1	99	98	89	0.94
EapH2	103	98	91	1.06
TSST-1	194	98	75	1.42
Local-global alignment using TSST-1 as a reference				
TSST-1	194	194	194	0
EAP2	98	194	76	1.45
EapH2	103	194	78	1.37
EapH1	99	194	75	1.45

exhibit notable structural divergence between these proteins (Fig. 2). The first of these regions lies between the  $\beta 1$  and  $\beta 2$  strands. In the case of EapH1 and EapH2, a short, additional  $\beta$  strand (denoted  $\beta 1'$  in Fig. 1, A and B) is found; this element is absent from EAP2, although strand  $\beta 2$  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  $\beta 5$  and  $\beta 6$ . In case of EapH1 and EapH2, the  $\beta 5$ – $\beta 6$  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  $\beta 5$  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  $\beta 1\beta 2$ , these values are 39.9, 39.6, and 30.2 Å<sup>2</sup>, whereas for  $\beta 5\beta 6$  they are 40.7, 31.1, and 20.6 Å<sup>2</sup>, where the mean temperature factors for the protein atoms are 30.2, 30.9, and 14.3 Å<sup>2</sup> 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

**FIG. 2. Least squares superposition of EapH1, EapH2, and EAP2 reveals regions of structural divergence.** The C $\alpha$  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  $\beta$ 1 $\beta$ 2 and  $\beta$ 5 $\beta$ 6.



the structural alignment algorithm DALI (35). Surprisingly, this revealed that EAP domains share significant structural homology with the secreted pyrogenic superantigen toxins TSST-1 (Protein Data Bank code 1AW7 (36);  $Z = 10.7$ ), and SEB (Protein Data Bank code 3SEB (37);  $Z = 8.2$ ), both from *S. aureus*, and SPE-C (Protein Data Bank code 1AN8 (38);  $Z = 9.1$ ) from *Streptococcus pyogenes*. Members of the superantigen family of proteins are composed of two structurally distinct domains of approximately equal size (39). The N-terminal domain consists primarily of a  $\beta$ -barrel structure that is similar to oligosaccharide/oligonucleotide-binding proteins (oligosaccharide-binding fold), whereas the C-terminal domain adopts a  $\beta$ -grasp motif that is structurally homologous to EAP domains (Fig. 3A) (39). When comparing the EAP2 and TSST-1 C-terminal domains, 76 of 98 C $\alpha$  atoms superimpose within 2.5 Å distance and with an r.m.s.d. of 1.45 Å (Fig. 3, B and C, and Table IV). This level of structural similarity is surprisingly high because a relationship between these classes of proteins cannot be detected by PSI-BLAST.

Comparison of these two protein structures revealed again that two major regions of structural divergence lay within the  $\beta$ 1 $\beta$ 2 and  $\beta$ 5 $\beta$ 6 linker regions. Of these, the most striking difference is found between the  $\beta$ 5 $\beta$ 6 regions, where a longer  $\beta$ 5 strand and more protruding loop conformation is seen in TSST-1<sub>CTD</sub> than in either EapH1, EapH2, or EAP2. Consistent with this observation, structure-based sequence alignment of several superantigen C-terminal domains with a group of EAP domains revealed that the  $\beta$ 5 $\beta$ 6 linker region is the major site of sequence variation between these classes of proteins (Fig. 3D). In fact, with the exception of TSST-1, each superantigen examined contains an insertion of at least 13 residues in this region relative to EAP domains. This region is involved in mediating interactions between the superantigen SEC2 and the T-cell receptor, supporting the idea that this variable region can be exploited to confer different binding specificities to EAP domains (Fig. 4).

#### 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  $\beta$ -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

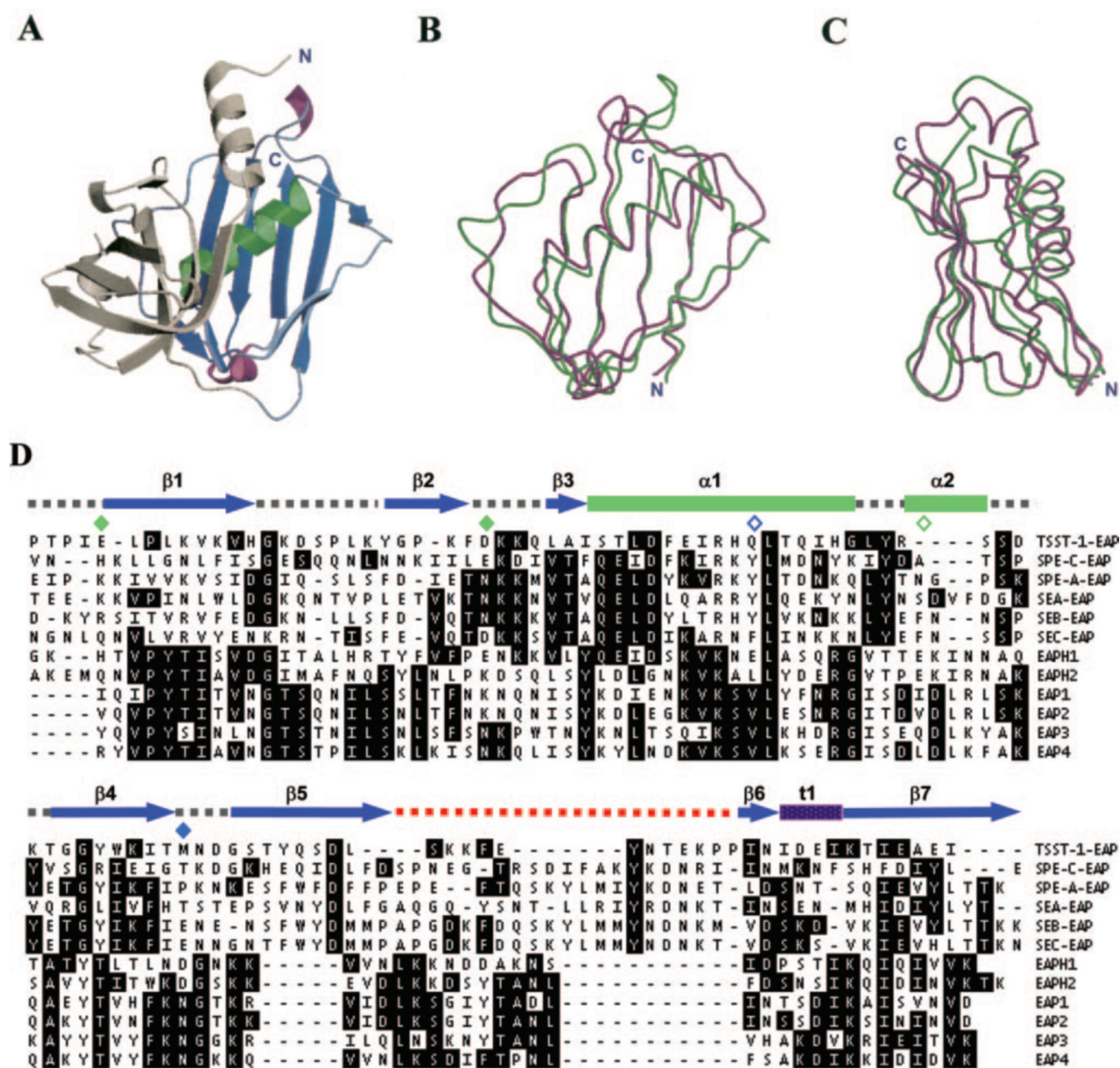
relationship is that EAP domains are a modular scaffold ideally suited to the formation of a wide variety of protein-protein interactions. Consistent with this hypothesis, the structural classification of proteins (SCOP) database indicates that the ability to participate in protein-protein interactions is a general feature of  $\beta$ -grasp proteins (40) ([scop.bic.nus.edu.sg/data/scop.b.e.bi.html](http://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.<sup>3</sup> 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 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  $\beta$ 5 $\beta$ 6-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  $\beta$ -chain revealed that residues lying within the

<sup>3</sup> B. V. Geisbrecht and D. J. Leahy, unpublished observations.





**FIG. 3. EAP domains share unexpected structural homology to the C-terminal domain of bacterial superantigens.** The structures of EapH1, EapH2, and EAP2 were analyzed for unexpected structural homology with structures deposited in the Protein Data Bank using the program DALI (35) (see “Results” and “Discussion”). **A**, a representative structure of the closest homolog, TSST-1 from *S. aureus*, is shown (36). The N-terminal, oligonucleotide/oligosaccharide-binding domain is depicted in gray scale, whereas the EAP-like domain is displayed in color as above. **B** and **C**, the  $\alpha$  backbone traces of EAP2 (purple) and the EAP-like domain from TSST-1 (green) were superimposed by least squares fitting using the program O (30) and presented at both 0 (**B**) and 90° (**C**) views as in Figs. 1 and 2. The protein termini are labeled accordingly. **D**, the sequences of the EAP-like, C-terminal domains from several superantigens were aligned with those of selected EAP domains using the DNASTAR program suite. Areas of conserved secondary structure are colored green ( $\alpha$ -helix), blue ( $\beta$ -sheet), and purple ( $3_{10}$ -helix) and shown above the appropriate residues. The  $\beta 5\beta 6$ -interstrand region, which is the main site of sequence and structural variation within this protein family, is denoted with a dashed red line. Zinc-coordinating residues in EapH1 are marked with colored diamonds, with the residues comprising site 1 in green and site 2 in blue. Side chains donated by symmetry related molecules are depicted as clear diamonds.

$\beta 5\beta 6$ -interstrand region are critical in forming this complex (1). Second, residues affecting the TCR  $\beta$ -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  $\beta 5\beta 6$ -interstrand region in toxin function (45). During the course of the studies presented here, we observed that the  $\beta 5\beta 6$ -interstrand region is the main area of sequence and structural divergence within this family of secreted virulence factors and toxins. This suggests that the  $\beta 5\beta 6$ -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  $\beta$ -subunit of the major histocompatibility class II proteins. Specifically, residues 3–33 in the mature major histocompatibility class II protein  $\beta$  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

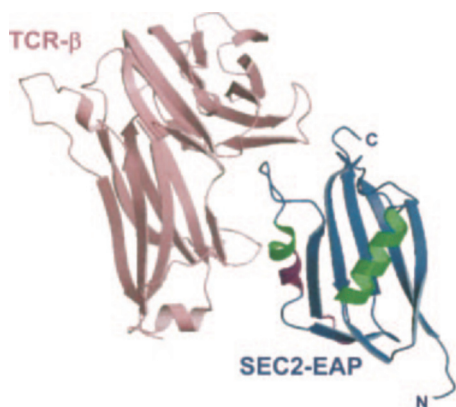


FIG. 4. **Structural analysis of superantigen complexes implicates the  $\beta 5\beta 6$ -interstrand region as a main determinant of specificity in EAP domain-mediated protein-protein interactions.** The structure of superantigen SEC2 complexed with the  $\beta$ -chain of a mammalian T-cell receptor (TCR- $\beta$ ) 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.

PSI-BLAST when a single EAP domain is used as a search query. To test whether the similarity between these regions was structurally significant, the  $\alpha$  atoms of the residues in question from the major histocompatibility class II protein molecule (HLA-DR1) were superimposed upon the corresponding residues of the EAP domains. This alignment revealed that the central region of the polypeptide chains were highly similar, both of which are comprised of two anti-parallel  $\beta$ -strands joined by a short turn to form a  $\beta$ -hairpin structure. The remainder of the polypeptide chains share little similarity, however, and the calculated r.m.s.d. of the  $\alpha$  superposition in each case is greater than 7.21 Å. This strictly confined region of weak structural similarity is insufficient to support a conclusion of a divergent evolutionary relationship and suggests that any resemblance is incidental.

While solving the EapH1 structure, two unexpected peaks of high electron density were observed at the contacts between neighboring molecules in the crystal lattice. Several lines of evidence, including the bond lengths, tetrahedral geometry, coordinating residues, and presence of zinc ions in the crystallization buffer, suggested that these sites represented structural zinc ions (Fig. 1A). Both heavy atom sites appear well ordered in the final EapH1 model ( $B_{\text{Zn-1}} = 27.8 \text{ Å}^2$ ,  $B_{\text{Zn-2}} = 42.3 \text{ Å}^2$ , where  $B_{\text{structure}} = 32.2 \text{ Å}^2$ ), a result that is consistent with the correct identification of these metal sites.

Zinc ions have been shown to mediate the formation of superantigen complexes in both crystal lattices and solution, and zinc appears to be required for the biological activity of these toxins (39, 46–49). Despite the many similarities between superantigens and EAP domains, zinc does not appear to be serving a conserved physiological role in EAP domain function. First, the zinc-coordinating residues in the EapH1 structure are not conserved in any other of the EAP domains examined (Fig. 3D). Not surprisingly, structural zinc ions were not observed in the EapH2 or EAP2 structure, nor were divalent cations required for crystallization of either of these proteins. Second, although each molecule of EapH1 binds two zinc ions in the structure, EapH1 does not exhibit metal-dependent multimerization under physiologically relevant zinc concentrations ( $[\text{Zn}^{2+}] \leq 1 \text{ mM}$ ; data not shown). Finally, no reports describe a requirement for zinc or other divalent cations in EAP domain structure or function. Thus, although it appears that metal binding is not a conserved feature of EAP domains, the possibility that divalent cations might contribute to the biochemical properties of specific EAP do-

main proteins, such as EapH1, cannot be ruled out at this time.

**Acknowledgments**—We are grateful to the High-Throughput Crystallization Laboratory of the Hauptmann-Woodward Institute (Buffalo, NY) for providing the initial crystals of EAP2. We also thank K. Ramyar, S. Bouyain, R. Porecha, and J. McLellan for assistance with x-ray data collection, R. Abramowitz and X. Yang of Beamline X4-A at the National Synchrotron Light Source for technical assistance during MAD data collection, and M. Bolbat of Bio-CARS Beamline 14BM-C at the Advanced Photon Source for additional native data collection.

## REFERENCES

- 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
- Lowy, F. D. (1998) *N. Engl. J. Med.* **339**, 520–532
- Kronvall, G., and Jonsson, K. (1999) *J. Mol. Recogn.* **12**, 1–7
- Patti, J. M., Allen, B. L., McGavin, M. J., and Hook, M. (1994) *Annu. Rev. Microbiol.* **48**, 585–617
- Flock, J.-I. (1999) *Mol. Med. Today* **5**, 532–537
- Yousif, Y., Mertz, A., Batsford, S. R., and Vogt, A. (1991) *Zentralbl. Bacteriol.* **21**, (suppl.) 168–169
- McGavin, M. H., Krajewska-Pietrasik, D., Ryden, C., and Hook, M. (1993) *Infect. Immun.* **61**, 2479–2485
- Boden, M. K., and Flock, J.-I. (1992) *Microb. Pathog.* **12**, 289–298
- Palma, M., Haggag, A., and Flock, J.-I. (1999) *J. Bacteriol.* **181**, 2840–2845
- Chavakis, T., Hussain, M., Kanse, S. M., Peters, G., Bretzel, R. G., Flock, J.-I., Herrmann, M., and Preissner, K. T. (2002) *Nat. Med.* **8**, 687–693
- Flock, M., and Flock, J.-I. (2001) *J. Bacteriol.* **183**, 3999–4003
- Kreikemeyer, B., McDevitt, D., and Podbielski, A. (2002) *Int. J. Med. Microbiol.* **292**, 283–295
- Hussain, M., Haggag, A., Heilmann, C., Peters, G., Flock, J.-I., and Herrmann, M. (2002) *Infect. Immun.* **70**, 2933–2940
- Haggag, A., Hussain, M., Lonnies, H., Herrmann, M., Norrby-Teglund, A., and Flock, J.-I. (2003) *Infect. Immun.* **71**, 2310–2317
- Jonsson, K., McDevitt, D., Homonylo, M., Patti, J. M., and Hook, M. (1995) *J. Biol. Chem.* **270**, 21457–21460
- Harraghy, N., Hussain, M., Haggag, A., Chavakis, T., Sinha, B., Herrmann, M., and Flock, J.-I. (2003) *Microbiology* **149**, 2701–2707
- Hussain, M., Becker, K., von Eiff, C., Peters, G., and Herrmann, M. (2001) *Clin. Diag. Lab. Immunol.* **8**, 1271–1276
- Marrack, P., and Kappler, J. (1990) *Science* **248**, 685–686
- Herman, A., Kappler, J. W., Marrack, P., and Pullen, A. M. (1991) *Ann. Rev. Immunol.* **9**, 745–772
- Marmur, J. (1961) *J. Mol. Biol.* **3**, 208–218
- Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577–2637
- Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P., and Hendrickson, W. A. (1994) *Proteins* **19**, 48–54
- Geisbrecht, B. V., Dowd, K. A., Barfield, R. W., Longo, P. A., and Leahy, D. J. (2003) *J. Biol. Chem.* **278**, 32561–32568
- Geisbrecht, B. V., Zhang, D., Schulz, H., and Gould, S. J. (1999) *J. Biol. Chem.* **274**, 21797–21803
- Beneken, J., Tu, J. C., Xiao, B., Nuriya, M., Yuan, J. P., Worley, P. F., and Leahy, D. J. (2000) *Neuron* **26**, 143–154
- Otwinski, Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S., ed) pp. 56–62, Daresbury Laboratory, Warrington, UK
- Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 849–861
- Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 965–972
- Terwilliger, T. C. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59**, 34–44
- Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **47**, 110–119
- Brunker, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905–921
- Collaborative Computational Crystallography Project 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30**, 1022–1025
- Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138
- Earhart, C. A., Mitchell, D. T., Murray, D. L., Pinheiro, D. M., Matsumura, M., Schlievert, P. M., and Ohlendorf, D. H. (1998) *Biochemistry* **37**, 7194–7202
- Papageorgiou, A. C., Tranter, H. S., and Acharya, K. R. (1998) *J. Mol. Biol.* **277**, 61–79
- Roussel, A., Anderson, B. F., Baker, H. M., Fraser, J. D., and Baker, E. N. (1997) *Nat. Struct. Biol.* **4**, 635–643
- Papageorgiou, A. C., and Acharya, K. R. (1997) *Structure* **5**, 991–996
- Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.* **247**, 536–540
- Rost, B., and Sander, C. (1993) *J. Mol. Biol.* **232**, 584–599
- Jones, D. T. (1999) *J. Mol. Biol.* **292**, 195–202
- Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996) *Cell* **84**, 155–164
- Irwin, M. J., Hudson, K. R., Fraser, J. D., and Gascoigne, N. R. J. (1992) *Nature* **359**, 841–843
- Acharya, K. R., Passalacqua, E. F., Jones, E. Y., Harlos, K., Stuart, D. I., Brehm, R. D., and Tranter, H. S. (1994) *Nature* **367**, 94–97

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., Schlievert, 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., Damdoui, 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