Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Ab initio structure determination of the lantibiotic mersacidin

Thomas R. Schneider,* Jörg Kärcher, Ehmke Pohl,† Paolo Lubini and George M. Sheldrick

University of Göttingen, Department of Structural Chemistry, Tammannstrasse 4, 37077 Göttingen, Germany

† Present address: European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany.

Correspondence e-mail: trs@shelx.uni-ac.gwdg.de

Received 17 December 1999
Accepted 7 March 2000

PDB Reference: mersacidin, 1qow.

The crystal structure of mersacidin, a potential novel antibiotic against methicillin- and vancomycin-resistant *Staphylococcus aureus* strains, has been determined by *ab initio* methods. Despite all crystals being merohedrally twinned, an accurate structural model with an R value of 13.4% has been obtained at atomic resolution. With six molecules in the asymmetric unit and no atom heavier than sulfur, the structure corresponds to a protein of 120 amino acids and is the largest approximately equal-atom unknown structure solved by direct methods. In the crystal, the molecule assumes a compact fold different from that found by NMR in solution. Comparison of the NCS-related molecules reveals regions of variable flexibility. The region highly homologous to the related antibiotic actagardine is very rigid and possibly defines an essential building block of this class of new antibacterial substances.

1. Introduction

Pathogenic bacteria are developing resistance to currently available antibiotics at an alarming rate. A particular threat is posed by methicillin-resistant *S. aureus* (MRSA) strains in hospitals, against which the glycopeptide antibiotic vancomycin is the only and last line of defence (Hawkey, 1998; Waldvogel, 1999). Several clinical cases of intermediate vancomycin resistance in *S. aureus* have been reported in recent years (Hiramatsu *et al.*, 1997; Smith *et al.*, 1999) and it has been shown that the genetic information related to vancomycin resistance in enterococci can be transferred to *S. aureus in vitro* (Noble *et al.*, 1992). A broader emergence of bacterial strains resistant against all available antibiotics, including vancomycin, may thus only be a question of time. There is an urgent need to identify and characterize new antimicrobial substances aimed at features of pathogenic cell metabolism not targeted by any therapeutics currently in use in order to be one step ahead of the pathogenic bacteria.

Lantibiotics (lanthionine-containing antibiotics), a group of bactericidal peptides that are produced by and mainly act against Gram-positive bacteria, are promising candidates for such novel antimicrobial drugs (Sahl & Bierbaum, 1998). The common chemical feature of lantibiotics is the presence of intramolecular rings formed by various modifications of the non-standard thioether amino acid lanthionine (Fig. 1). Bacteria capable of producing lanthionines not only possess the genetic information for the peptide itself but entire clusters of genes encoding the cellular machinery for post-translational modification, transport and activation of these compounds and last, but not least, for the protection of the producing cell against its own antibacterial product.

Mersacidin is produced on the ribosome as a 68 amino-acid long peptide (SWISSPROT accession code P43683) including a 20 amino-acid C-terminal propeptide. Threonine and serine residues of the propeptide part are dehydrated by specific enzymes leading to 2,3-didehydroalanine (Dha) and 2,3-didehydrobutyrine (Dhb). This is followed by the stereospecific addition of thiol groups of cysteine residues to the double bonds of Dhb to form β-methyllanthionines (Jung, 1991). The C-terminal residue in mersacidin has been degraded to a 2,3-dehydrolanthionine (Fig. 1). Proteolytic cleavage of the 48 amino-acid leader sequence results in a strongly cross-linked peptide with a molecular weight of 1824 Da.

The biological activity of mersacidin has been studied extensively by a number of groups. The molecule interferes with the biosynthesis of the cell wall in the transglycosylation phase by binding to the final monomeric peptidoglycan precursor undecaprenyl-pyrophosphoryl-MurNAc-(pentapeptide)-GlcNAc (lipid II; Brötz *et al.*, 1997, 1998). The exact mode of interaction is unknown, but it has been shown that it does not involve the target site of vancomycin, the C-terminal D-Ala-D-Ala moiety of the lipid intermediate (Brötz *et al.*, 1995), and so far no cross-resistance for vancomycin and mersacidin has been found. Nevertheless, the *in vivo* activities of mersacidin and vancomycin against MRSA strains are comparable (Chatterjee, Chatterjee, Jani *et al.*, 1992), making mersacidin a possible alternative to vancomycin.

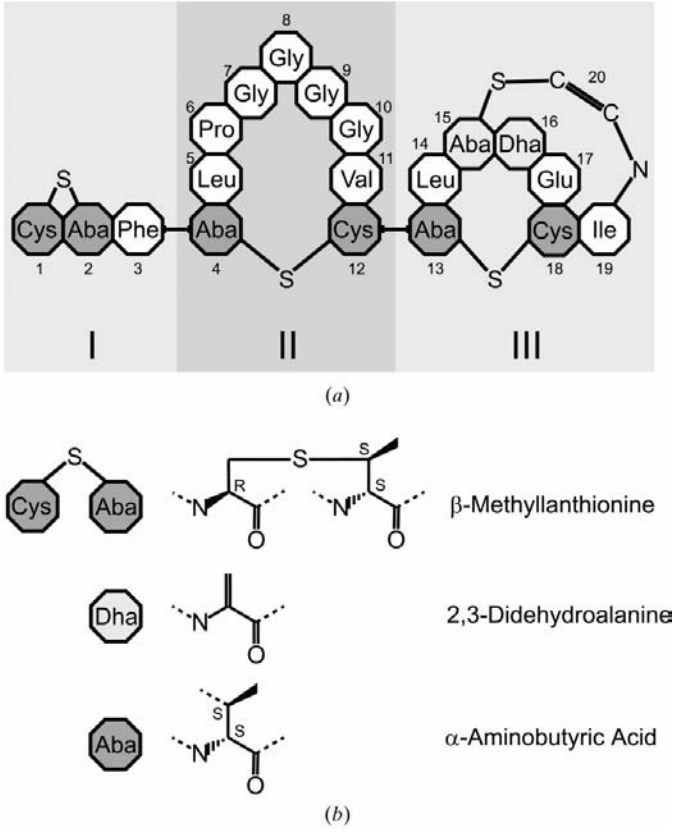

Figure 1
(*a*) Primary structure and thioether-bridging pattern of mersacidin. The molecule can be divided into three domains (I, II, III) based on the thioether links. Dha stands for dehydroalanine and Aba for α-aminobutyric acid. (*b*) Structure formulas of the non-standard amino acids.

This paper reports the first crystal structure determination of a lantibiotic. At first sight, with 125 non-H atoms per molecule, the structure determination of mersacidin should be a typical large 'small-molecule' problem. However, the crystals formed by mersacidin contain six antibiotic molecules, corresponding to 750 atoms plus solvent, per asymmetric unit. In addition, all crystals showed significant signs of merohedral twinning, a phenomenon that is a serious obstacle to most structure-determination methods and that is receiving increased attention by the macromolecular crystallographic community (Yeates & Fam, 1999; Breyer *et al.*, 1999; Dumas *et al.*, 1999; Müller *et al.*, 1999; Lücke *et al.*, 1998).

As in many antibiotics and small proteins, the extremely tight packing of the molecules in the crystal prevents the preparation of isomorphous heavy-atom derivatives; artificial incorporation of anomalous scatterers (such as selenomethionine residues) is not possible for most compounds from natural sources. Various strategies were tried to solve the structure by molecular replacement using search models derived from the solution structure (Prasch *et al.*, 1997), none giving a satisfactory solution. After all attempts using traditional protein crystallographic methods failed, *ab initio* methods, *i.e.* methods that find an initial phase set based exclusively on the experimentally measured diffraction intensities, became the last resort.

Although the quality of the data was severely degraded by the application of a detwinning procedure, an initial phase set was found by the program *SHELXD* (also known as '*Halfbaked*'; Sheldrick, 1998), which determines phases by real/reciprocal-space iteration (Miller *et al.*, 1993, 1994). Mersacidin represents the largest unknown approximately equal-atom structure solved so far by such *ab initio* methods. Subsequent refinement against the original data led to an accurate structural model, despite the merohedral twinning.

The crystal structure contains six crystallographically independent molecules that not only provide a static picture of the molecule but also, by comparison of the different copies, allow an assessment of its conformational flexibility. Substantial differences in overall conformation were found for all six molecules relative to the solution structure, thus in retrospect explaining the failure to solve the structure by molecular replacement. The observed flexibility may play a decisive role in defining the function and the specificity of the molecule.

2. Materials and methods

2.1. Purification and crystallization

Mersacidin was isolated from *Bacillus* species HIL Y 85,54728 and purified as described in Chatterjee, Chatterjee, Lad *et al.* (1992). The lyophilyzed powder was dissolved to saturation in methanol. Crystals grew at room temperature by slow evaporation of methanol from a 1:1 mixture of the saturated solution with benzene and reached maximum

Table 1
Data statistics for mersacidin.

Redundancy and completeness are given for space group $P3_2$. $R_{int} = \sum|I - \langle I\rangle|/\sum|\langle I\rangle|$, where $\langle I\rangle$ is the mean intensity of a set of equivalent reflections. The highest resolution shell contains data in the resolution range 1.06–1.15 Å.

Wavelength (Å)	0.865
Temperature (K)	100
Resolution (Å)	∞–1.06
Reflections collected	164063
Reflections unique/possible	33469/33485
Unit-cell parameters (Å)	$a = b = 46.09$, $c = 31.02$
Average redundancy	4.9
Completeness (%)	100.0
$\langle I/\sigma(I)\rangle$ (overall/highest resolution shell)	14.9/3.4
$\langle\lvert E^2 - 1\rvert\rangle$	0.649
R_{int} $P3_2$ (%)	4.9
R_{int} $P3_221$ (%)	19.5
R_{int} $P3_212$ (%)	44.3

dimensions of 0.3 × 0.3 × 0.3 mm in 10 d. The crystals exhibit hexagonal shape and belong to space group $P3_2$, with unit-cell parameters $a = b = 46.09$, $c = 31.02$ Å, $\alpha = \beta = 90$, $\gamma = 120°$, and are merohedrally twinned to varying degrees. We were not able to control the degree of merohedral twinning in a reproducible way. For data collection, crystals were transferred to a perfluorinated polyether oil (Hope, 1988; Kottke & Stalke, 1993), attached to a glass fibre and flash-cooled to 100 K in a stream of gaseous nitrogen.

2.2. Data collection

Data were collected on beamline BW7B at EMBL c/o DESY (Silfhout & Hermes, 1995) using a MAR180 image-plate detector. The data were collected at 100 K in three passes to limiting resolutions of 1.05, 1.5 and 3.0 Å, respectively, at a wavelength of 0.865 Å. Diffraction intensities were integrated and merged in Laue group $\bar{3}$ using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) and particular care was taken to use a consistent indexing scheme in all data processing. Subsequent merging was performed with *XPREP* (Bruker Analytical X-ray Systems). Data-collection statistics are given in Table 1.

2.3. Detwinning

For a merohedrally twinned crystal consisting of two domains A and B, pairs of reflections hkl and $h'k'l'$ are twin-related: each diffraction intensity represents a weighted average of reflections from the domains A and B,

$$I_{hkl} = (1 - k)I_{A,hkl} + kI_{B,h'k'l'},$$
$$I_{h'k'l'} = kI_{A,hkl} + (1 - k)I_{B,h'k'l'},$$

where k is the relative contribution of domain B and the relation between the indices hkl and $h'k'l'$ is determined by the twinning operation (or twin law). For space group $P3_2$, the twinning operation emulating space group $P3_221$ (as indicated by the lower R_{int} in comparison with space group $P3_212$; Table 2) is a 180° rotation about the short diagonal in the ab plane, so reflections hkl and $h'k'l'$, where $h' = k$, $k' = h$ and $l' = -l$, are twin-related. Based on this twin law, the true intensities from both domains can be estimated *via*

$$I_{A,hkl} = [(1 - k)I_{hkl} - kI_{h'k'l'}]/(1 - 2k),$$
$$I_{B,h'k'l'} = [(1 - k)I_{h'k'l'} - kI_{hkl}]/(1 - 2k),$$

provided the twinning ratio k is not equal to 0.5. Using this formalism, the original data were deconvoluted into the contributions from two domains, employing different values for the twinning ratio k. For the resulting data, $\langle|E^2 - 1|\rangle$ (where E is a normalized structure factor) and the number of negative intensities produced were monitored (Table 2). A twinning fraction of $k = 0.2$ gave the best agreement between the calculated and the theoretical value for $\langle|E^2 - 1|\rangle$ while giving an acceptable number of negative intensities and so was employed to generate 'detwinned' data.

Table 2
Number of negative reflections ($I < 0$) and $\langle|E^2 - 1|\rangle$ for data obtained by deconvoluting the original data using different values for the twinning ratio k.

k	$I < 0$	$\langle\lvert E^2 - 1\rvert\rangle$
0.0	657	0.643
0.1	892	0.675
0.2	1927	0.735
0.3	5180	0.838

2.4. Structure solution

The data obtained from the detwinning operation were input into the *ab initio* program *SHELXD* (Sheldrick, 1998). Starting from many different sets of random atoms, the program first attempted to find 32 atoms by dual-space recycling, as at the time it was thought that there might be 32 S atoms in the asymmetric unit (the correct number turned out to be 24). The solutions with the highest correlation coefficients (CC; Fujinaga & Read, 1987) between observed and calculated normalized structure factors E_{obs} and E_{calc} were then expanded to the full structure by seven cycles of 'peak-list optimization' (Sheldrick & Gould, 1995), an E-Fourier recycling procedure. The 435th random starting set gave a CC of 24.5% before peak-list optimization and 71.0% after and was clearly a correct solution; the highest CC values for false solutions were 14.9 and 57.1%, respectively.

2.5. Refinement

All refinement was performed against the original intensities employing the method of Pratt *et al.* (1971) and Jameson *et al.* (1982) as implemented in *SHELXL*97. Here, the effect of merohedral twinning is described by a fixed twin matrix [in this case: (0 1 0, 1 0 0, 0 0 −1)] and a refinable twinning ratio k. In the very first round of refinement, k refined to a value of 0.34 and then gradually decreased towards its final value of 0.24.

The initial model was built by assigning atoms to peak positions found by *SHELXD* using the program *XP* (Bruker

Analytical X-ray Systems) and comprised 658 of the 750 mersacidin atoms, giving a conventional R value of 34.2%. For refinement, the geometries of standard amino acids were restrained to standard target values (Engh & Huber, 1991). For the non-standard amino acids, the presence of six copies of the molecule in the asymmetric unit was exploited to generate 'similarity restraints', *i.e.* restraints that drive chemically equivalent distances to the same target value, whereby the target value itself is allowed to vary. Using the restraints described above, the distortions of the initial model built against the detwinned data could be resolved and the R value dropped from 34.2 to 27.3 (DFIX, SADI in Fig. 2). Subsequent cycles included model building based on $3F_{obs} - 2F_{calc}$ and $2F_{obs} - 2F_{calc}$ difference Fourier maps using *XtalView* (McRee, 1999).

Although it is possible to *refine* directly against the observed intensities for the overlapping twin components – indeed, refinement against 'detwinned' data is incorrect, because the observations become highly correlated – it is unfortunately necessary to introduce an approximation in order to calculate electron-density *maps*. The problem is that the difference between the observed intensity and the sum of the calculated intensities of the twin components needs to be partitioned between the components, but it is not clear what the best way of partitioning is. The 97–2 version of *SHELXL* used for this project assigns partial F^2_{obs} values to the individual components in the same ratio as their F^2_{calc} values. Maps are then calculated exactly as for untwinned structures using the F_{calc} for the 'prime' component (normally the component with the largest scattering volume) and the 'partial' F_{obs} for this component obtained by the partitioning.

The refinement proceeded smoothly until the two most mobile regions (Gly7–Gly10 of molecules C and D) had to be modelled. At this stage, automatic placement of dummy atoms using the program *SHELXWAT* (Sheldrick & Schneider, 1997) based on $F_{obs} - F_{calc}$ difference peaks in a fashion similar to the program *ARP* (Lamzin & Wilson, 1997) greatly improved the electron-density maps: all peaks higher than 4.5 r.m.s. of the map and further than 1.3 Å away from existing atoms were interpreted as dummy atoms using oxygen scattering factors with a starting B value of 20 Å^2; dummy atoms whose B value exceeded 32 Å^2 after five cycles of conjugate-gradient B-value and coordinate refinement were eliminated. After ten sets of five cycles of this procedure, both missing tetraglycine regions could be positioned and several other parts of the model could be substantially improved, lowering the R value to 19.2%. Inclusion of restrained anisotropic replacement parameters gave an R value of 15.8%. The model was completed by adding H atoms on calculated positions and was finally subjected to a cycle of unrestrained full-matrix least-squares minimization in order to derive estimated standard deviations. One matrix inversion for 33 449 observables and 7438 parameters used 112 Mbytes of memory and required 5.8 h of CPU time on a Pentium II PC running at 450 MHz under the Linux operating system. Estimated standard deviations of atomic positions and hydrogen-bond lengths were determined using the full covariance matrix as implemented in *SHELXL*97. For non-standard 1–2 and 1–3 distances in the lanthionine groups, the errors given are calculated as the standard deviation of the distribution of the respective distances in the final model.

The final model has an R value of 13.4% for reflections with $F > 4\sigma(F)$ (14.1% for all reflections) and contains all 750 mersacidin atoms, 58 water molecules and nine methanol molecules. The model is in good agreement with the imposed restraints on geometry and ADPs (Table 3).

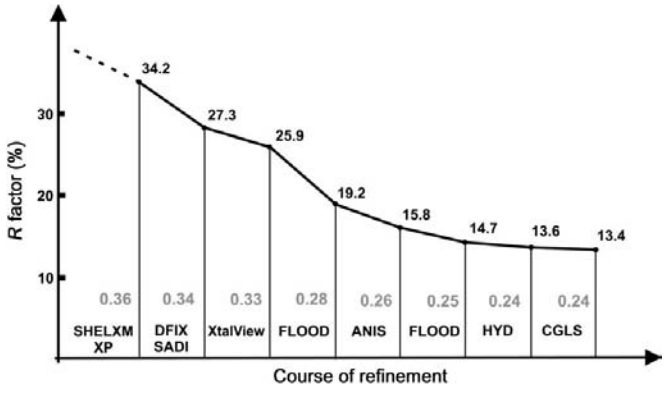

Figure 2
Course of refinement for mersacidin. R values are given for reflections with $F > 4\sigma(F)$. Values of the twinning parameter k at the end of each cycle are shown in grey. DFIX and SADI stand for the use of Engh & Huber and similarity restraints, respectively, XtalView for model building using *XtalView*, FLOOD for density modification by placement of dummy atoms, ANIS for inclusion of anisotropic displacement parameters, HYD for inclusion of H atoms and CGLS for final conjugate-gradient minimization.

Table 3
Statistics of the final model.

In the final model all non-H atoms were refined with individual restrained anisotropic displacement parameters.

Refinement statistics	
Resolution range (Å)	39.9–1.06
No. of reflections used in refinement	33449
No. of parameters refined	7438
Crystallographic R factor [$F > 4\sigma(F)$/all reflections] (%)	13.4/14.1
Atoms in model	
Mersacidin (non-H)	750
Hydrogen	719
Water (excluding H atoms)	58
Methanol (excluding H atoms)	18
R.m.s. deviation from restraint target values	
Bond lengths (Å)	0.013
Angle distances (Å)	0.034
Similar distances (no target values) (Å)	0.017
Distances from restraint planes (Å)	0.027
Zero chiral volumes (Å^3)	0.069
Non-zero chiral volumes (Å^3)	0.061
Anti-bumping distance restraints (Å)	0.016
Rigid-bond ADP components (Å^2)	0.005
Similar ADP components (Å^2)	0.052
Approximately isotropic ADPs (Å^2)	0.068
Average B factors for non-H atoms	
Mersacidin main chain (Å^2)	15.5
Mersacidin side chain (Å^2)	18.9
Water (Å^2)	20.9
Methanol (Å^2)	19.9

2.6. Structure analysis

Hydrogen-bonding patterns were analyzed using *XP* (Bruker Analytical X-ray Systems). Accessible surfaces were calculated using the program *SURFACE* in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) with a probe radius of 1.4 Å. The mean accessible surface area of a monomer was determined to be 1695 ± 36 Å^2; the mean surface area of a monomer within the hexamer is 1049 ± 39 Å^2. The NCS-related molecules were compared using error-scaled difference distance matrices calculated with the program *ESCET* (Schneider, 2000). Least-squares superposition of molecules was performed using *SHELXPRO* (Sheldrick & Schneider, 1997). Figures were prepared with *MOLSCRIPT* (Kraulis, 1991), *BOBSCRIPT* (Esnouf, 1997) and *Raster3D* (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

3. Results and discussion

3.1. Detection and overcoming of merohedral twinning

The data set extends to atomic resolution and is of good quality, as reflected in the completeness of the data and the high signal-to-noise ratio at the high-resolution limit. Merging statistics and systematic absences indicate the trigonal space group $P3_2$ (or the enantiomorphic $P3_1$), one of the space groups that allow merohedral twinning to occur. Merohedrally twinned crystals consist of domains whose lattices superimpose exactly, thus producing a perfect diffraction pattern although the sample is not a single crystal (Giacovazzo, 1992; Yeates, 1997). The usual way to detect merohedral twinning is by inspection of the intensity statistics. These revealed two characteristic warning signs of twinning (Herbst-Irmer & Sheldrick, 1998). Firstly, the variance of the normalized structure factors, $\langle|E^2 - 1|\rangle$, has an abnormally low value of 0.649 (the expected value for non-centrosymmetric space groups is 0.736), indicating a 'washed-out' intensity distribution in which, owing to the averaging of reflections from different domains, extremely low and high intensities are systematically underrepresented. Secondly, the $R_{\rm int}$ value (Table 1) in the higher symmetry trigonal space group $P3_221$ (19.8%) is substantially lower than the $R_{\rm int}$ for the alternative possible higher symmetry space group $P3_212$ (44.3%).

For a merohedral twin, each measured intensity $I_{\rm tot}$ is the superposition of two intensities $I_{A,hkl}$ and $I_{B,h'k'l'}$ from the two domains A and B, which have different but fixed relative orientations such that the indices hkl and $h'k'l'$ are related by a mathematically exact twin law. In the case of the mersacidin crystals, the better merging in space group $P3_221$ suggests a twin law corresponding to an additional twofold axis along the short diagonal in the ab plane of the trigonal cell. Using this twin law and the relative size of the different domains, untwinned intensities $I_{A,hkl}$ and $I_{B,hkl}$ can be extracted from $I_{\rm tot}$ (see §2). However, as the relative size of the domains is not known at the outset, it has to be determined for example by assuming that the resulting 'detwinned' data set should follow normal intensity statistics as closely as possible. Here, a twinning fraction of 0.2 was found to give the best detwinned data in terms of $\langle|E^2 - 1|\rangle$ being close to its theoretically expected value and an acceptable number of negative intensities (Table 2).

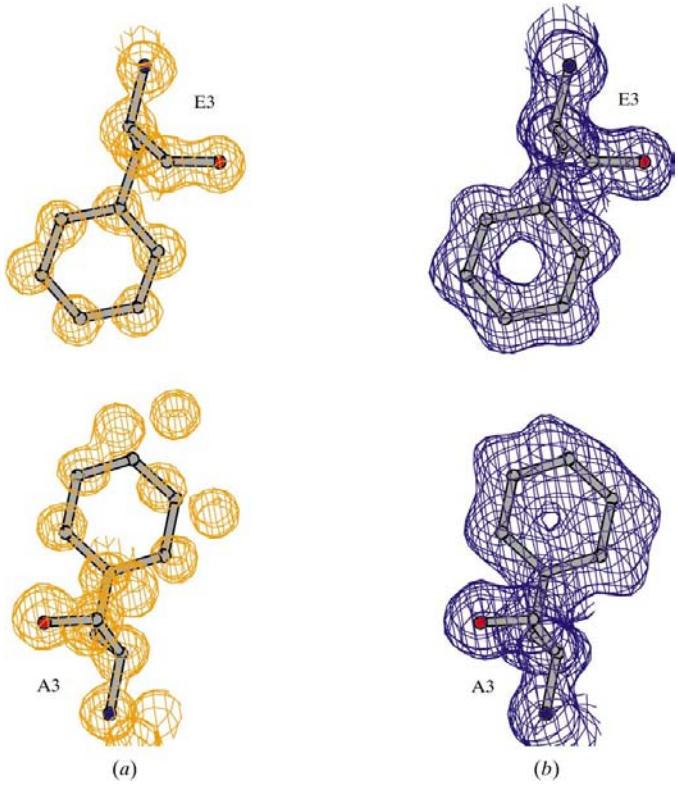

Figure 3
Electron densities representing the result of *ab initio* phasing and the final model. In both parts of the figure, coordinates of the final model are shown for residues PheA3 and PheE3. (*a*) shows a Fourier synthesis based on structure factors F_c and phases φ_c calculated from atoms found by *SHELXD* contoured at 2.0σ. (*b*) shows a $2mF_o - DF_c\,\sigma_A$-weighted difference electron density based on structure factors and phases from the final model contoured at 1.2σ.

3.2. *Ab initio* structure solution

Phases for the detwinned intensity data were found using the program *SHELXD* (Sheldrick, 1998). *SHELXD* requires as the only input the number of atoms of different type to find, the unit cell and crystal symmetry and a file containing diffraction intensities. As output, the program produces a list of atomic positions sorted by peak height. After about 3 d CPU time on a Pentium Pro computer running at 200 MHz, positions for 658 out of 750 non-H atoms (corresponding to six molecules in the asymmetric unit) were found. These positions were taken to built the initial model based on the known sequence (Chatterjee, Chatterjee, Jani *et al.*, 1992) by assigning atoms to positions without needing to refer to electron-density maps. Fig. 3(*a*) shows a representation of these positions in relation to the final model. It should be noted that this procedure is – in contrast to most methods of determining protein structures – entirely free of 'model bias' because the

atomic positions were obtained without making any stereochemical assumptions.

3.3. Refinement against merohedrally twinned data

Refinement was carried out against the original (*i.e.* not detwinned) data taking into account the above twin law by the method of Pratt *et al.* (1971) and Jameson *et al.* (1982) as implemented in *SHELXL*97 (Sheldrick & Schneider, 1997; Herbst-Irmer & Sheldrick, 1998). For a given twin law, this method introduces only one additional parameter, the twinning fraction k, into the refinement process. Refining the value of k simultaneously with the 'normal' parameters of a crystallographic refinement such as coordinates and displacement parameters allows one to find a model that is consistent with the original measured X-ray intensities. This approach has the advantage that no additional systematic errors are introduced into the data by the detwinning procedure. The effect of such errors can be seen in Fig. 3: placing atoms into the positions found based on the detwinned data will result in a very distorted model. In fact, although the overall connectivity was as expected from the sequence and the solution structure, the initial model was severely distorted in several regions, hampering smooth refinement. Use of tight stereochemical restraints helped to overcome this problem. For standard amino acids, Engh & Huber (1991) geometries were employed. For the lanthionine moieties as well as the Dha residues, no suitable model structures could be found in the Cambridge Structural Database (Allen, 1998), so the high degree of non-crystallographic symmetry was exploited by restraining chemically equivalent 1–2 and 1–3 distances to be similar with a given standard deviation but without specifying a fixed target value.

After defining local NCS (non-crystallographic symmetry) restraints for the non-standard amino acids, the refinement proceeded smoothly until the most mobile regions of the structure had to be modelled. Even when all data to 1.06 Å were used, no continous electron density was found for residues Gly7–Gly10 in two of the six molecules (*C* and *D*). This was not surprising as difference Fourier maps are expected to be inaccurate in intermediate stages of a refinement against twinned data because of the difficulty of partitioning the differences between the observed and calculated intensities between the different twin domains. Nevertheless, automatic placement and removal of dummy atoms based on difference Fourier maps and *B*-factor refinement in a manner similar to the *ARP* procedure (Lamzin & Wilson, 1997) led to a significant improvement of electron-density maps and allowed the completion of even the most mobile regions of the model. The final model contains all H and non-H atoms for the six mersacidin molecules, 58 water molecules and nine methanol molecules. The final crystallographic R factor is 13.4% for $F > 4\sigma(F)$ with a refined ratio of the twin domains of 0.24:0.76, a value close to the value estimated in the initial detwinning procedure.

Although the refinement statistics indicate that the atomic model is of high quality, electron-density maps calculated for the final model appear to be somewhat worse than expected for a crystal structure with data collected to 1.06 Å resolution from an untwinned crystal (Fig. 3*b*). This reflects the uncertainty in partitioning the difference structure factors discussed above.

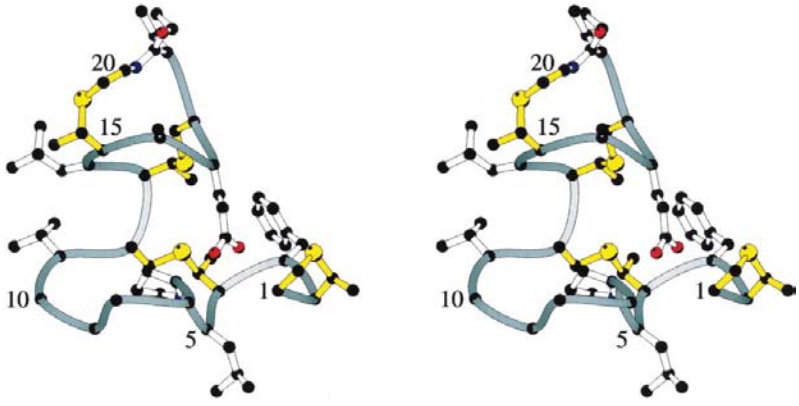

Figure 4
Stereoview of mersacidin. The polypeptide backbone is shown in dark grey, with transitions between domains in light grey. Thioether bridges are shown in yellow. The view is along a least-squares plane through all 69 atoms of domains I and II except the side chain of Leu5. The mean distance of atoms from this plane is between 0.96 (molecule *B*) and 1.05 Å (molecule *E*).

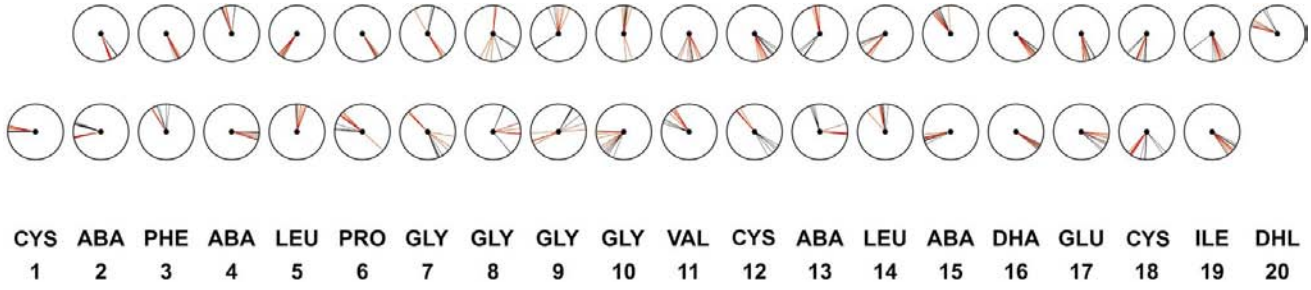

Figure 5
Schematic representation of backbone dihedral angles Φ (top row) and Ψ (bottom row) for the five NMR structures (black) and the six molecules in the crystallographic asymmetric unit (red). Dihedral angles are shown in anticlockwise direction, with 3 o'clock corresponding to 0°.

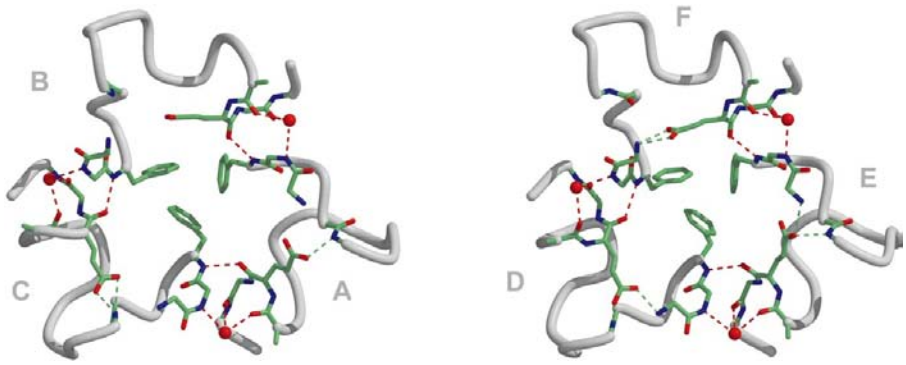

Figure 6
Schematic representation of important hydrogen bonds. Molecules are labelled *A*–*E*. The polypeptide backbone is shown as a light grey tube; where hydrogen bonds are formed, individual atoms are shown. For Glu17 and Phe3, side-chain atoms are depicted. Water molecules are shown in red. Intramolecular hydrogen bonds are shown as light green, intermolecular hydrogen bonds as red dashed lines.

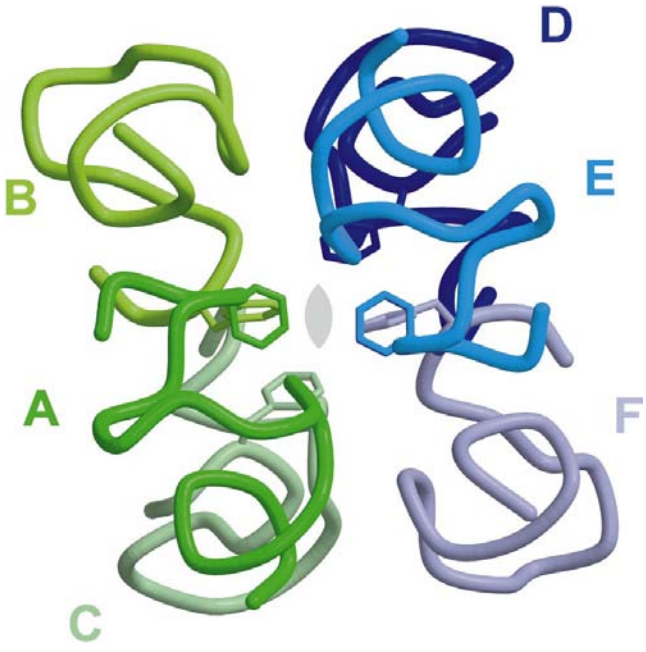

Figure 7
Arrangement of the two trimers (molecules *A*, *B* and *C* in green; molecules *D*, *E* and *F* in blue). View along the diagonal of the *ab* plane in the crystal. The pseudo-twofold axis is indicated in grey.

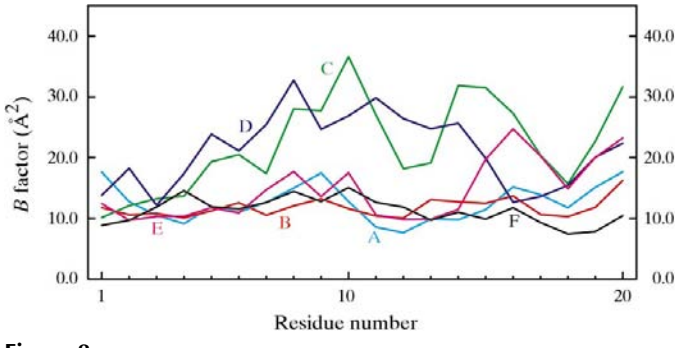

Figure 8
Average *B* factor for backbone (N—CA—C) atoms *versus* residue number for the six molecules *A*–*F* in mersacidin.

3.4. Description of the molecule

All six molecules present in the crystallographic asymmetric unit show the same compact fold mostly defined by the thioether-bridging pattern (Fig. 1). Domains I (residues 1–3) and II (residues 4–12) form a remarkably flat almost rectangular base of approximately 14 × 17 Å in size that is held rigid by two thioether links and hydrogen bonds in the β-hairpin-type turn comprising residues 1–5 (Fig. 4). Domain III (residues 13–20) is attached to the rim of this base and bends back onto domain I and II to position the carboxylate group of Glu17 so that it can interact with the N-terminal portion of the molecule.

The crystal and solution structures are very similar with respect to the thioether-bridging pattern and the backbone conformation within domains I and III. However, domain II contains a stretch of four consecutive glycine residues (Gly7–Gly10) found to have a large conformational variability in both the NMR and the X-ray ensembles. Local differences in this region between the NMR and the X-ray structure lead to a different backbone fold of domain II and, in combination with a variation in the dihedral angles between residues Ala12 and Aba13 that link domains II and III, result in a substantially different overall conformation (Figs. 4, 5 and 8).

3.5. Oligomeric state and structural base of twinning

The six molecules in the asymmetric unit are arranged with approximate 32 (D_3) point-group symmetry (Figs. 6 and 7) and can be described as a dimer of trimers. The trimers are held together by a very regular intermolecular hydrogen-bonding pattern linking the amide N atom of Phe3 and the carbonyl O atom of Glu17 of adjacent molecules and reinforced by water-molecule mediated interactions (Fig. 6). The formation of the hexamer, in which about 40% of the surface of the hypothetical free monomer becomes buried, is favoured by the resulting central hydrophobic cluster of six phenylalanine side chains.

The only two electrostatic charges on mersacidin at physiological pH are located on the carboxylate group of Glu17 and on the N-terminal amino group. They are on the surface of the hexamer for all six molecules. In the conformation of mersacidin found in the crystal, the charges approximately neutralize each other (Table 4), resulting in a non-charged particle and explaining the insolubility of pure mersacidin in water. However, if mersacidin is dissolved in pyridine, a water-soluble salt can be formed (Chatterjee, Chatterjee, Lad *et al.*, 1992) by addition of a low-concentration aqueous solution of KOH. The resulting salt is soluble in water and shows the same biological activity (Chatterjee, Chatterjee, Lad *et al.*, 1992) as mersacidin itself.

Table 4
Intramolecular hydrogen bonds between the carboxy group of Glu17 and domain I.

Distances between non-H atoms are given in Å, with standard uncertainties in parentheses. The distances in square brackets correspond to weak or absent hydrogen bonds.

	Glu17···Ala1		Glu17···Gly7	
Mol.	Glu17 OE1···Ala1 N	Glu17 OE2···Ala1 N	Glu17 OE1···Gly7 N	Glu17 OE2···Gly7 N
A	[3.7 (1)]	[4.4 (1)]	[3.73 (9)]	2.61 (6)
B	[4.96 (5)]	[5.19 (7)]	[4.91 (7)]	[4.81 (9)]
C	[4.0 (2)]	[5.1 (2)]	2.8 (1)	2.5 (2)
D	2.61 (7)	[3.6 (1)]	[3.4 (1)]	2.56 (9)
E	[3.45 (5)]	2.87 (7)	[3.91 (6)]	2.56 (5)
F	2.83 (4)	2.88 (4)	[6.08 (4)]	[5.47 (5)]

Table 5
1–2 and 1–3 distances for β-methyllanthione as derived from the refined structure of mersacidin.

Standard uncertainties in the last digits are given in parentheses.

Atoms	Distance (Å)
C^{α}(Cys)–C^{β}(Cys)	1.53 (6)
C^{α}(Cys)–S^{γ}(Aba)	2.74 (4)
C^{β}(Cys)–S^{γ}(Cys)	1.80 (4)
C^{β}(Cys)–C^{β}(Aba)	2.80 (5)
C^{α}(Aba)–C^{β}(Aba)	1.53 (5)
Cα(Aba)–S^{γ}(Cys)	2.79 (4)
C^{α}(Aba)–C^{γ}(Aba)	2.52 (6)
C^{β}(Aba)–S^{γ}(Cys)	1.81 (4)
C^{β}(Aba)–C^{γ}(Aba)	1.53 (5)
C^{γ}(Aba)–S^{γ}(Cys)	2.75 (4)

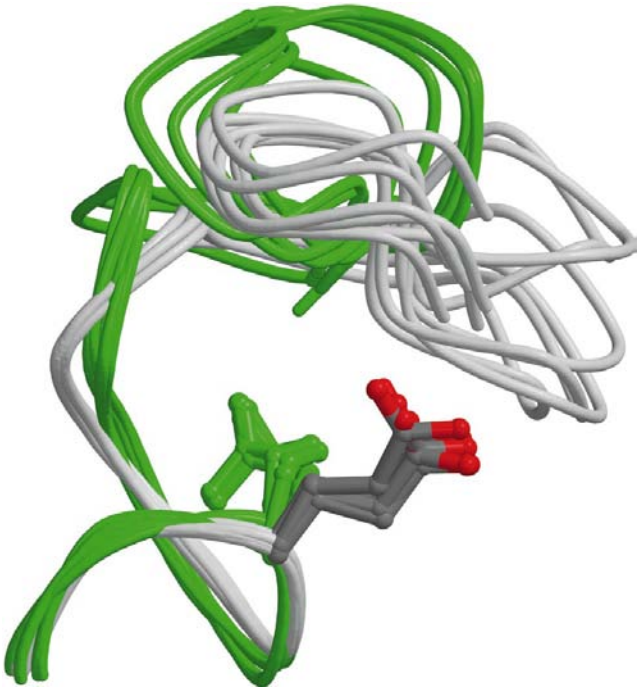

Figure 9
Least-squares superposition based on the CA atoms of domain III for molecules *A*–*F*. The backbone is shown in light grey, the side chain Glu17 in dark grey and red. For the NMR ensemble, the polypeptide backbones and the side chains of Glu17 are shown in green after least-squares superposition of CA atoms of residues 12–20 with corresponding atoms of molecule *A* from the X-ray structure.

The trimers within the hexamers are related by a pseudo-twofold axis. In the crystal, this axis runs along the short diagonal of the *ab* plane and, if fulfilled exactly, would give rise to the higher symmetry space group $P3_221$. This pseudo-symmetry and the fact that there are no strong interactions between the hexamer and other hexamers related by crystallographic symmetry operations allows the addition of hexamers to a growing crystal in the 'wrong' orientation (*i.e.* rotated by 180°) with almost no disturbance of the resulting crystal lattice. Such anomalous growth of a crystal will lead to 'a regular aggregate consisting of crystals of the same species joined together in some definite mutual orientation' (Giacovazzo, 1992) where the reciprocal lattices of the different twin domains superimpose exactly. Thus, the non-crystallographic symmetry of the hexamer provides an explanation of the macroscopically observed phenomenon of merohedral twinning at the microscopic level.

3.6. Geometry of lanthionines

The high data quality and the presence of multiple copies of the non-standard amino acid β-methyllanthionine allow the determination of their stereochemistry. Averaging 1–2 and 1–3 distances for the 18 copies of β-methyllanthionine in the final model gives relatively accurate and precise estimates for bond lengths and angles (Table 5) that can be used for future modelling, NMR and crystallographic studies on related compounds.

3.7. Conformational variability

Being of small size, a defined flexibility may be essential for mersacidin to achieve high specifity. With average *B* factors around 15 Å^2, four of the six mersacidin molecules are relatively rigid (Fig. 8), while the other two (*C* and *D*, twin-related) have significantly higher average *B* factors, indicating higher flexibility. Furthermore, the presence of six copies of mersacidin in the asymmetric unit provides a unique opportunity to assess the flexibility of mersacidin by analyzing the conformational variability between six structural models that are all based on the same experimental data.

A simple least-squares superposition of molecules is often not an appropriate approach to compare molecules, as the choice of atoms for superposition is subjective and strongly influences the result. Additionally, the different degrees of coordinate reliability in different molecules or regions of molecules are not properly accounted for in currently used superposition algorithms. The definition of a conformationally invariant region of a molecule, taking into account the level of precision of the individual atoms, can be conveniently and objectively performed using error-scaled difference distance matrices as described in the accompanying paper (Schneider, 2000). This analysis revealed that domain III is the most rigid part of the structure and that domains I and II can move

relative to domain III. A superposition based on the CA atoms of residues belonging to domain III shows almost identical backbone conformations (Fig. 9), whereas domains I and II are more flexible entities in themselves and additionally can move relative to domain III. In fact, domain III was also found to be the least flexible part of the molecule in the NMR investigation (Prasch *et al.*, 1997).

Owing to the floppiness of domains I and II, a large variation can be seen in the hydrogen-bonding pattern between the carboxylate group of Glu17 and the N-terminal part of the molecule. Although the carboxylate group is always found in a very similar conformation relative to domain III, the details of interaction with its partners Ala1 and Gly7 are different in each molecule (Fig. 6 and Table 4): very tight interaction on one hand (molecule *D*) contrasts with rather loose interaction on the other (molecule *B*). Its large variation suggests that this polar interaction is rather loose and not the main determinant of the fold of the molecule.

The region between Val11 and Cys18 is, with the exception of Dha16, identical in sequence to region Val5–Cys12 of the related lantibiotic actagardine (Zimmermann & Jung, 1997). Actagardine shows a similar mode of action, thus suggesting a functional role for domain III. It is tempting to speculate that Glu17, which in the hexameric form of mersacidin is on the surface of the oligomer, is lingering in a resting position interacting with the opposite charge on the N-terminus until an appropriate ligand becomes available. Flapping out of the carboxylate to interact with the ligand will change intramolecular and intermolecular interactions (Fig. 6), leading to a break-up of the hexamer. Once in the monomeric form, mersacidin would have much more conformational flexibility, allowing specific interaction with its ligand.

We are grateful to László Vértesy for providing the material and helpful discussions, Zbyszek Dauter and Regine Herbst-Irmer for discussions and Ludger Häming for help during data collection. We thank C. Griesinger and coworkers for providing coordinates for the solution structure. This work was supported by the Fonds der Chemischen Industrie (EP, GMS). PL thanks the Swiss Bundesamt for Bildung und Wissenschaft for a postdoctoral grant (BBW 94.0162).

References

Allen, F. H. (1998). *Acta Cryst.* A**54**, 758–771.
Bacon, D. J. & Anderson, W. F. (1988). *J. Mol. Graph.* **6**, 219–220.
Breyer, W. A., Kingston, R. L., Anderson, B. F. & Baker, E. N. (1999). *Acta Cryst.* D**55**, 129–138.
Brötz, H., Bierbaum, G., Leopold, K., Reynolds, P. E. & Sahl, H.-G. (1998). *Antimicrob. Agents Chemother.* **42**, 154–160.
Brötz, H., Bierbaum, G., Markus, A., Molitor, E. & Sahl, H.-G. (1995). *Antimicrob. Agents Chemother.* **39**, 714–719.
Brötz, H., Bierbaum, G., Reynolds, P. E. & Sahl, H.-G. (1997). *Eur. J. Biochem.* **246**, 193–199.
Chatterjee, S., Chatterjee, D. K., Jani, R. H., Blumbach, J., Ganguli, B. N., Kiesel, N., Limbert, M. & Seibert, G. (1992). *J. Antibiot.* **45**, 839–845.
Chatterjee, S., Chatterjee, S., Lad, S. J., Phansalkar, M. S., Rupp, R. H., Ganguli, B. N., Fehlhaber, H.-W. & Kogler, H. J. (1992). *J. Antibiot.* **45**, 832–838.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dumas, P., Ennifar, E. & Walter, P. (1999). *Acta Cryst.* D**55**, 1179–1187.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Fujinaga, M. & Read, R. (1987). *J. Appl. Cryst.* **20**, 517–521.
Giacovazzo, C. (1992). *IUCr Texts on Crystallography 2. Fundamentals of Crystallography*. Oxford Science Publications.
Hawkey, P. M. (1998). *Lancet*, **351**, 1298–1299.
Herbst-Irmer, R. & Sheldrick, G. M. (1998). *Acta Cryst.* B**54**, 443–449.
Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Ogurim, T. & Tenover, F. C. (1997). *J. Antimicrob. Chemother.* **40**, 135–136.
Hope, H. (1988). *Acta Cryst.* B**44**, 22–26.
Jameson, G. B., Schneider, R., Dubler, E. & Oswald, H. R. (1982). *Acta Cryst.* B**38**, 3016–3020.
Jung, G. (1991). *Angew. Chem. Int. Ed. Engl.* **30**, 1051–1068.
Kottke, T. & Stalke, D. (1993). *J. Appl. Cryst.* **26**, 615–619.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Lamzin, V. L. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.
Lücke, H., Richter, H.-T. & Lanyi, J. K. (1998). *Science*, **280**, 1934–1937.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A (1993). *Science*, **259**, 1430–1433.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Müller, U., Muller, I. A., Herbst-Irmer, R., Sprinzl, M. & Heinemann, U. (1999). *Acta Cryst.* D**55**, 1405–1413.
Noble, W. C., Virani, Z. & Cree, R. (1992). *FEMS Microbiol. Lett.* **72**, 195–198.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–325.
Prasch, T., Naumann, Th., Markert, R. L. M., Sattler, M., Schubert, W., Schaal, S., Bauch, M., Kogler, H. & Griesinger, C. (1997). *Eur. J. Biochem.* **244**, 501–512.
Pratt, C. S., Coyle, B. A. & Ibers, J. A. (1971). *J. Chem. Soc. A*, pp. 2146–2151.
Sahl, H.-G. & Bierbaum, G. (1998). *Annu. Rev. Microbiol.* **52**, 41–79.
Schneider, T. R. (2000). *Acta Cryst.* D**56**, 714–721.
Sheldrick, G. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Silfhout, R. G. & Hermes, C. (1995). *Rev. Sci. Instrum.* **66**, 1818–1820.
Smith, T. L., Pearson, M. L., Wilcox, K. R., Cruz, C., Lancaster, M. V., Robinson-Dunn, B., Tenover, F. C., Zervos, M. J., Band, J. D., White, E. & Jarvis, W. R. (1999). *N. Engl. J. Med.* **340**, 493–501.
Waldvogel, F. A. (1999). *N. Engl. J. Med.* **340**, 556–557.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–360.
Yeates, T. O. & Fam, B. C. (1999). *Structure*, **7**, R25–R29.
Zimmermann, N. & Jung, G. (1997). *Eur. J. Biochem.* **246**, 809–819.