A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity

Caitríona A. DENNIS*†[1], Hortense VIDELER*§, Richard A. PAUPTIT†, Russell WALLIS‡, Richard JAMES*, Geoffrey R. MOORE§ and Colin KLEANTHOUS*[2]

*School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, U.K., †Protein Structure Laboratory, Zeneca Pharmaceuticals, Mereside, Alderley Park, Macclesfield SK10 4TG, U.K., ‡Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and §School of Chemical Sciences, University of East Anglia, Norwich, NR4 7TJ, U.K.

We report the first detailed comparison of two immunity proteins which, in conjunction with recent protein engineering data, begins to explain how these structurally similar proteins are able to bind and inhibit the endonuclease domain of colicin E9 (E9 DNase) with affinities that differ by 12 orders of magnitude. In the present work, we have determined the X-ray structure of the *Escherichia coli* colicin E7 immunity protein Im7 to 2.0 Å resolution by molecular replacement, using as a trial model the recently determined NMR solution structure of Im9. Whereas the two proteins adopt similar four-helix structures, subtle structural differences, in particular involving a conserved tyrosine residue critical for E9 DNase binding, and the identity of key residues in the specificity helix, lie at the heart of their markedly different ability to bind the E9 DNase. Two other crystal structures were reported recently for Im7; in one, Im7 was a monomer and was very similar to the structure reported here, whereas in the other it was a dimer to which functional significance was assigned. Since this previous work suggested that Im7 could exist either as a monomer or a dimer, we used analytical ultracentrifugation to investigate this question further. Under a variety of solution conditions, we found that Im7 only ever exists in solution as a monomer, even up to protein concentrations of 15 mg/ml, casting doubt on the functional significance of the crystallographically observed dimer. This work provides a structural framework with which we can understand immunity-protein specificity, and in addition we believe it to be the first successfully refined crystal structure solved by molecular replacement using an NMR trial model with less than 100 % sequence identity.

INTRODUCTION

Bacteria co-existing under competitive conditions are able to secrete toxins, known as bacteriocins, which can kill their competitors. Bacteriocins encompass a wide spectrum of chemical substances, ranging from small-molecule antibiotics, lantibiotics and microcins, to *Lactobacillus* bacteriocins and large bacterial proteins, such as the pyocin family from *Pseudomonas aeruginosa*, and the colicin family from strains of enterobacteria [1]. Colicins are plasmid-encoded bacteriocins, synthesized by *Escherichia coli* and *Shigella senni*, which have structural domains that perform different functions [2]; the N-terminal domain is implicated in translocation across the membrane of the target cell, while recognition of the target cell is achieved by a central domain that binds to a specific extracellular surface receptor. The C-terminal domain houses the toxic activity of the protein.

Colicins can be identified broadly according to the translocation pathway they use to penetrate cells. Group A colicins, comprising colicins A, E1–E9, M, N and K, enter cells by the translocation pathway encoded by the *tol* Q,R,A,B gene cluster, whereas group B colicins, comprising colicins B, D, Ia and Ib, use the translocation pathway of the Ton B system (reviewed by James et al. [1]). Colicins are also classified according to the extracellular receptors they use to initiate cell death. The group E colicins, for example, recognize the BtuB protein of *E. coli* cells, which is an outer-membrane receptor specific for vitamin B_{12} transport [3].

The cytotoxic activities of colicins, all of which are carried out by the C-terminal domain of the toxin, generally fall into three classes. The largest class is the pore-forming proteins, which includes colicins A, B, E1, Ia, Ib, K and N. These act via the destruction of the membrane potential, resulting in cell death [4–7]. The two other classes that have been identified within the E group of colicins act enzymically. RNases, comprising colicins E3, E5 and E6, catalyse the specific cleavage of the 16S rRNA from the 30S subunit of ribosomes, resulting in the inhibition of protein synthesis [8,9]. The DNase colicins ColE2, ColE7, ColE8 and ColE9 act by the non-specific endonucleolytic cleavage of both single-stranded and double-stranded DNA [10,11], and it is this latter class of toxin that has been the focus of our work.

Nine E-group colicins have been identified (ColE1–ColE9) by immunity testing [12,13]. The basis for this test is the requirement of a colicin-producing cell to protect itself against its own colicin. Hence, E-colicin plasmids contain an *imm* gene that encodes a specific immunity protein (identified by the prefix Im), which binds to the toxic domain of its cognate colicin and inhibits its cytotoxic activity [11,14,15]. The nuclease-type immunity proteins (9.5 kDa) bind very tightly to the colicin (61 kDa), ensuring that the toxin is inactive inside the producing cell [16]. The 70 kDa heterodimeric complex is then released into the environment with the assistance of a third protein encoded on the plasmid, the lysis protein [17]. Following receptor binding by the toxin, it is thought that the bound immunity protein is dissociated, by an unknown mechanism, as the colicin begins to penetrate a bacterium.

We have been using the DNase colicin–immunity-protein system to address questions of specificity in protein–protein interactions within families of structurally similar proteins, an area of molecular recognition that is still poorly understood. The

Abbreviation used: E9 DNase, the isolated 15 kDa endonuclease domain of ColE9.

[1] Present address, School of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, U.K.
[2] To whom correspondence should be addressed (e-mail c.kleanthous@uea.ac.uk).

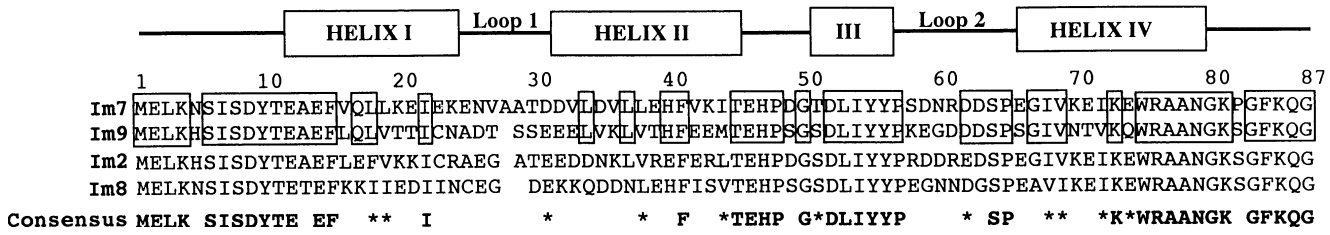

Figure 1 Sequence alignment of all the DNase E colicin immunity proteins

Identical residues between Im7 and Im9 are boxed. * Indicates a conservative substitution.

four DNase colicins are almost identical in the N-terminal regions of the protein involved in receptor binding and translocation, but are only 80 % identical in their C-terminal DNase domains. Hence, specific immunity proteins are needed for each toxin and the sequence identity in this family of proteins is $\approx 50\,\%$ ([18]; Figure 1).

The interaction between E-colicin DNase domains and immunity proteins can be studied both *in vivo* and *in vitro*. In naturally occurring bacterial isolates, colicin-immunity protein recognition is highly specific, since any given immunity protein will not, in general, provide protection for a non-cognate DNase toxin. However, non-cognate cross-reactivities can be readily detected by overexpressing immunity genes in bacteria. Thus, although Im9 is the cognate immunity protein for ColE9, biological protection towards ColE9 by Im8 and Im2 was detected by Wallis et al. [19] when these were overexpressed in bacterial cells, the latter giving stronger protection. In contrast, Im7 did not provide any biological protection towards ColE9.

These biological cross-reactivities are mirrored by binding affinities *in vitro*. We have determined the dissociation constant (K_d) for the cognate endonuclease domain of colicin E9 (E9 DNase)–immunity-protein complex as 9.3×10^{-17} M in buffers of low ionic strength, making this one of the highest affinity complexes yet reported [16]. Interestingly, the affinities of the 61 kDa toxin and the isolated 15 kDa E9 DNase domain, which can be expressed separately from the rest of the toxin [20,21], are identical in complex with Im9. We have also found that all the non-cognate immunity proteins can bind and inhibit the activity of the E9 DNase, but with much weaker affinities: 10^{-4} M for Im7, 10^{-6} M for Im8 and 10^{-8} M for Im2 [19].

Structures of colicin DNases and immunity proteins in both the free and bound states are needed in order to explain protein–protein interaction specificity at the molecular level. At present, there is no structural information available for any colicin DNase and little is known of the location of the active site, nor how the DNase activity is inhibited by the binding of an immunity protein. Nevertheless, putative active-site residues in the E9 DNase have been identified by random mutagenesis [21], and specificity-determining residues have been identified by site-directed mutagenesis [22]. Rather more is known about the immunity proteins. We have solved the solution structure of Im9 and probed the surface of contact with its 15 kDa cognate E9 DNase domain by high-field NMR spectroscopy [23–25]. The data show that Im9 is a distorted, four-helix protein and that helices II and III are likely to form the DNase binding site. The structural data have been the basis for homologue-scanning mutagenesis experiments, which identified helix II, one of several variable regions in this family of proteins (Figure 1), as the main determinant of immunity specificity [26].

Two crystal structures of Im7 have recently been reported, the co-ordinates of which were not available at the time this work was carried out, which show Im7 to have a similar fold to that of Im9. In one of these, Im7 occurs in a monomeric form [27], whereas in the other it is dimeric [28]. The Im7 dimer was reported to be of functional significance in terms of the auto-regulation of Im7 expression. In the present paper, we have addressed the issue of the oligomeric structure of Im7 by analytical ultracentrifugation. We have also undertaken the crystal structure of Im7 by molecular replacement, using Im9 as a trial model, and discuss the ColE9 specificity differences of the two immunity proteins through a structural comparison.

MATERIALS AND METHODS

Protein purification and characterization

Im7 was purified and quantificated as described by Wallis et al. [16,19]. The molecular mass of the purified protein was determined by electrospray MS, as described by Garinot-Schneider et al. [21] and Li et al. [26].

Analytical ultracentrifugation

All experiments were carried out in a Beckman Optima XL-A analytical ultracentrifuge equipped with absorbance optics using an An60Ti rotor. Im7 samples (110 μl) in 50 mM potassium phosphate buffer (pH 7.0), containing 0–500 mM NaCl and 1 mM dithiothreitol, were placed in the sample cells of an Epon charcoal-filled six-channel centrepiece. The reference cells were loaded with aliquots of the buffer (125 μl). Experiments were performed at 30000 and 32000 revs./min at 20 °C. The partial specific volume of Im7 (0.731 ml/g) was calculated from the amino acid composition using the method of Cohn and Edsall [29].

Equilibrium data were collected at 295–300 nm in step-scan mode using a separation of 0.001 cm. Five readings were averaged for each scan and five separate scans averaged. Readings were taken at 4 h intervals until no difference could be detected between consecutive scans. The equilibrium distributions from three different loading concentrations (two-fold dilutions up to 9 mg/ml) were fitted simultaneously to a single non-ideal species model using the Nonlin curve-fit algorithm [30]. Data were also analysed by calculation of the apparent weight-average molecular mass at concentrations throughout the sample cells using software supplied with the ultracentrifuge.

Crystallization and X-ray data collection

Initial trials for crystallization were carried out using Hampton Research (Laguna Hills, CA, U.S.A.) Crystal screens I and II using the automated microbatch method implemented on an IMPAX robotic system (Douglas Instruments; London, U.K.).

Duplicate plates were set up so that one could be incubated at 20 °C and one at 4 °C. Hanging-drop vapour diffusion was also used as a complementary screen.

The screening experiments resulted in microcrystals around pH 4.6 with a variety of precipitants. Optimization, micro-seeding techniques and additive screens were unsuccessful at increasing crystal size, as were variations in protein concentration, temperature and precipitant. Crystals for Im7 were eventually obtained with one of the salt scans (dibasic ammonium phosphate) using the hanging-drop method. Conditions for crystal growth involved mixing a drop of 10 mg/ml protein with an equal volume of 55 % saturated di-ammonium phosphate in 200 mM Tris buffer at pH 7.2, and using 100 % saturated di-ammonium phosphate as a reservoir in the well. After 2 days incubation at 20 °C, thin plates measuring 0.5 mm × 0.5 mm × 0.02 mm were obtained. The crystal symmetry was consistent with space group I222 or $I2_12_12_1$ and had unit cell dimensions of $a = 45.1$, $b = 50.6$ and $c = 75.2$ Å, apparently isomorphous to the crystals used by Chak et al. [27].

Crystals of Im7 were mounted in thin-walled glass capillaries, diameter 0.5 mm (Hampton Research). The X-ray source used was an Enraf Nonius FR571 rotating anode, operating at 40 mV and 90 mA with a copper target, emitting a wavelength of 1.5418 Å (CuK_α). Data were collected on a 30 cm MAR Image plate (Xray Research, Hamburg, Germany). The detector was positioned at 150 mm from the centred crystal. X-ray exposure time per image was 10 min and corresponded to one degree of crystal rotation about the horizontal ϕ axis. A total 200 images were collected from one crystal. The crystal of Im7 showed diffraction to resolution 1.8 Å, with only a slight weakening of the high-resolution reflections towards the end of the collection. The images were processed using XDS software [31,32] and CCP4 programs [33].

Structure solution and refinement

The Im7 structure was determined by molecular-replacement methods, using a superimposed ensemble of the ten lowest-energy NMR structures of Im9 as a trial model [24]. Regions of main-chain flexibility encompassing the loop regions and the N- and C-termini (residues 1–5, 25–31, 56–62 and 82–86) were removed from the trial model and non-conserved side chains were replaced by alanine. Conserved side-chains that showed differing conformations in the ensemble were also truncated to alanine. The resultant trial model contained ≈ 60 % of the protein, but included all the secondary-structure elements (helical regions). The average root-mean-squares deviation for the back-bone of the 10 superimposed NMR structures was computed for the Im9 model, and this value (0.53 Å) was converted to an average atomic pseudo-temperature factor ($B = 20$ Å^2).

Molecular-replacement trials were carried out using the program AMORE [33–35]. Resolution ranges varying between 15 and 4.5 Å, and a radius of integration varying from 14 to 16 Å, were used, and translation functions were carried out in both possible space groups I222 and $I2_12_12_1$. Using a resolution range 15–4.5 Å and the space group I222, a correct translation solution was obtained. This was the highest peak but it did not stand out significantly above the rest; its correlation coefficient was only 2 % higher and its R-factor only marginally lower than the others. The rotation peak for which a correct translation-function solution was found ranked 57th in the list of peaks obtained from the rotation-function search. Rigid-body refinement of the translation search solution resulted in an increased correlation co-efficient of 56 %, 4 % higher than the next peak, and the R-factor decreased by 2 % to 48 %, 2 % lower than the next peak. The molecular replacement was not trivial, and use of an ensemble of structures as a trial model was critical for success in this case; trials carried out using a single NMR structure did not yield any answers, whether or not the flexible portions of the model were excised. The solution was validated through inspection of the crystal packing, which showed tight packing with reasonable contacts to symmetry-related positions and no steric violations. Calculation of a 2Fo-Fc electron-density map (contoured at 1 σ) showed good density for the helices as well as density for some of the side-chains that had been trimmed to alanines. The appearance of density for missing parts of the model was taken as absolute confirmation of a correct molecular-replacement solution and of the correct space group.

Whereas with the program AMORE the molecular replacement could be carried out using an ensemble of trial models, a single member of the ensemble (that of the lowest energy) was selected as a starting model for refinement. The starting model was incomplete due to the excised disordered portions, and also not as representative of the true solution as was the ensemble, hence the initial derived structure amplitudes and phases were not of high quality. Refinement was initiated using standard simulated annealing protocols of the program XPLOR, interspersed with manual rebuilding using the graphics program O [36]. The R-factor decreased from its initial value of 0.544 (Rfree = 0.501) to 0.432 (0.485) and 0.388 (0.478) after two rounds of simulated annealing using data between 8 and 3 Å, and to 0.365 (0.462) after a third round of simulated annealing using data between 8 and 2 Å. At this stage, some side-chain density corresponding to the correct sequence was visible and these side chains were rebuilt. Hereafter, simulated annealing was replaced by conventional positional refinement, but, whereas R could be reduced to 0.31, Rfree remained high (0.43), indicating over-fitting. Residues built into weak loop density were removed, and the maximum-likelihood refinement program REFMAC [33,37,38] was applied to the incomplete model. In 10 cycles, using data between 10 and 2 Å, Rfree dropped significantly to 0.33. Subsequently calculated electron-density maps were of better quality, and positive density could be seen for most of the loops. The loop structure was rebuilt and validated by another 10 cycles of REFMAC refinement, which now lowered R and Rfree to 0.262 and 0.301, respectively. Refinement with REFMAC was continued, allowing the rebuilding of all residues, except Gly87, for which no density could be observed, until R and Rfree were 0.212 and 0.265, respectively. At this stage, the highest peaks in the difference map corresponded to density of water molecules, and water structure was built into peaks higher than 3 standard deviations of the map using QUANTA software (MSI; San Diego, CA, U.S.A.). Positional refinement, temperature-factor refinement and adjustment of solvent molecules were continued until no further difference density could reasonably be ascribed to water and the refinement had converged with an R-factor of 17.8 % and an Rfree of 24.7 %. Analysis using the program PROCHECK [39] showed that 94.7 % (72 of 76) of residues have back-bone torsion angles in the most favoured regions of the Ramachandram plot, with 5.2 % (4 of 76) in additionally allowed regions (not counting 4 glycines, 4 prolines and 2 terminal residues). Co-ordinates and structure factors have been deposited in the Brookhaven protein databank; the accession code is 1ayi. Table 1 shows the data-processing statistics.

RESULTS AND DISCUSSION

Oligomeric structure

Chak, Yuan and co-workers [28] have described recently a novel autoregulatory role for Im7 (termed ImmE7 by them) in which

Table 1 Data-processing statistics

Cell parameters	$a = 45.1$ Å, $b = 50.6$ Å, $c = 75.2$ Å, $\alpha = \beta = \gamma = 90°$
Space group	I222 or $I2_12_12_1$
Completeness (overall)	93%
Completeness (highest shell 2.18–2.0 Å)	87%
R_{sym} (overall)	4.9%
R_{sym} (highest shell 2.18–2.0 Å)	15.6%
Total number of reflections	19772
Number of unique reflections	5828
Volume to mass ratio	2.2 Å^3/Da
Solvent content	41%

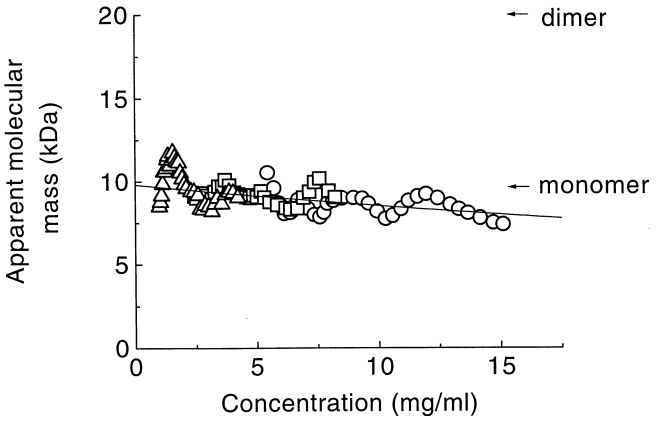

Figure 2 Equilibrium ultracentrifugation analysis of Im7

Shows the apparent weight-averaged molecular mass distribution with concentration for three samples of Im7 at 32000 rpm in 50 mM potassium phosphate buffer (pH 7.0), containing 200 mM NaCl and 1 mM dithiothreitol at 20 °C. The loading concentrations were approximately 9 mg/ml (○), 4.5 mg/ml (□) and 2.3 mg/ml (△). The calculated masses of the monomer and dimer are indicated by arrows. The data were fitted to a single non-ideal species model.

its dimerization creates an RNase-like active site specific for its own mRNA. They found that under similar crystallization conditions, two crystal forms of Im7 appeared and the structures for both were solved. The first reported structure was that of the monomer [27], which predominated in their crystallization experiments, whereas the second was that of the dimeric form. Crystals of the dimer could not be grown reproducibly [40]. The protein structures in the two crystal forms are very similar, except that the protein is seen to dimerize along helix II in the second form. It was argued that changes in the oligomeric structure of Im7 underpin its different functions; the monomer is a DNase inhibitor (specific for the toxin ColE7), whereas the dimer is a putative RNase. No solution experiments concerning the oligomeric structure of Im7 have yet been presented in the literature and so we sought to determine whether Im7 showed a tendency to dimerize in solution.

The molecular mass of expressed and purified Im7 was determined by electrospray MS as 9893.9 Da, which corresponds closely to that predicted from the sequence (9894.9 Da; results not shown). Equilibrium ultracentrifugation experiments were conducted at approximately the same pH as the subsequent crystallography, as described in the Materials and Methods section, and some of these data are presented in Figure 2. The weight-averaged molecular mass of Im7 was estimated to be 9800 ± 300 Da from two independent experiments in buffer containing 200 mM salt, in good agreement with the monomer molecular mass. No oligomerization was detected under any of the conditions examined (including 500 mM NaCl), at protein concentrations up to 15 mg/ml; Im7 crystals were obtained at 10 mg/ml in this study. There is a decrease, however, in the apparent weight-averaged molecular mass with concentration, reflecting non-ideal behaviour (Figure 2), most likely through charge-repulsion effects. Im7, in common with all the DNase immunity proteins, is highly acidic, with a calculated pI of 4.3. Consistent with this, the second virial coefficient B (the non-ideal term) was found to vary with salt concentration; estimated as 1.8×10^{-6} and 6.6×10^{-6} l·mol·g^{-2} in buffer containing 200 mM NaCl and no salt, respectively. Similar findings have been reported previously for Im9, which is also monomeric in solution [15].

Our results demonstrate that Im7 does not dimerize in solution. The formation of a dimer, reported by Hsieh et al. [28], may well be the result of crystal-lattice contacts rather than an intrinsic property of the protein, and there is certainly precedent for changes in oligomeric structure induced by the environment in a crystal. RNase A, for example, is a monomer in solution and in many crystal structures but can, under certain crystallization conditions, form dimers in a crystal [41]. In such cases, it is important to consider whether the amount of buried surface between monomers is typical of functional dimers. The amount of buried surface reported by Hsieh et al. [28] at the dimer interface, ≈ 870 Å^2, is signficantly smaller than most stable protein–protein complexes, which tend to bury > 1500 Å^2, but of the same order as 'crystallographic dimers' [42].

Use of NMR structures as trial models

Although difficult and time-consuming, the structure determination and refinement of Im7, starting with an NMR trial model of 60% sequence identity, was successful. The molecular replacement was not trivial, and use of an ensemble of structures as a trial model was critical for success in this case; trials carried out using a single NMR structure did not yield any answers, whether or not the flexible portions of the model were excised. To the best of our knowledge, published molecular-replacement success using NMR structures as trial models has been hitherto restricted to cases of 100% sequence identity, i.e. where the crystal structure has been determined previously by NMR [43–48]. It is also well documented that there are difficulties in refining molecular-replacement solutions using NMR trial models [49]. Taking such considerations into account, it would seem that the structure determination of Im7 must be close to the limit of the capabilities of molecular replacement. The difficulties arise from the fact that, in an NMR structure, much of the protein (loops, surface side chains) is modelled as flexible and offers limited information towards the conformation in the crystal, where loops and surface side chains more often have defined structure. In the case of Im7, this amounted to 40% of the structure. The remaining part of the NMR model is still not the best approximation for a crystal structure but, in the molecular replacement, increased overlap with the calculated and observed Patterson functions may be obtained by using an ensemble of NMR structures. The use of maximum-likelihood-refinement methodology (program REFMAC; [37,38]) in treating incomplete and inaccurate starting models also proved pivotal.

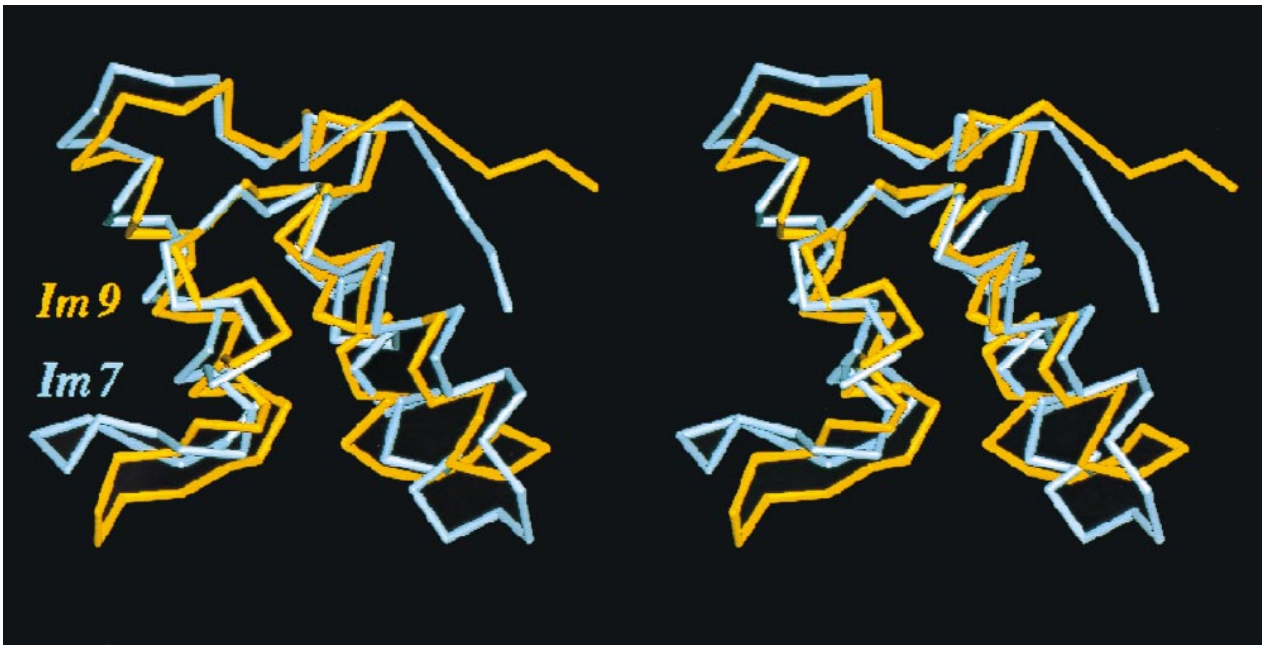

Figure 3 Stereo representation of Im7 determined by molecular replacement overlayed with the Im9 solution structure

The structures were superimposed by least-squares alignment of equivalent Cα atoms of the helices only. The N-terminus of each protein is on the right-hand side of the Figure. Helices II and III are at the front of Figure going from right to left.

Topology

The 87 residue Im7 protein is folded into an approximate four-helix bundle topology (Figure 3). It differs from a classical four-helix bundle in the length and orientation of the α-helices; in particular, helix III is very short, being terminated by Pro57. There is a turn of 3_{10}-helix between Ser6 and Tyr10. The four α-helices are labelled I–IV and summarized in Table 2. They are all amphipathic helices, containing polar side chains, which lie on the surface of the molecule, as well as hydrophobic side chains, which are buried in the centre of the molecule, in approximately equal numbers. The hydrophobic residues from each helix are oriented towards each other to form the hydrophobic core of the molecule. Loops I and II follow α-helices I and III, respectively. There is a short linker (Glu46–Gly50) separating helices II and III. Although the loops might be expected to be mobile from their variable conformations in the NMR ensemble, in the crystal structure they are ordered with reasonable temperature-factor values. The average B value for all non-hydrogen atoms in loop I (Glu25–Asp31) is 26.3 Å², whereas the average for the entire structure is 21.6 Å². Similarly, in loop II (Pro57–Pro65) the average temperature factor is 26.3 Å². In the crystal, loop I from one molecule packs against loop II from a neighbouring symmetry-related molecule, so it is conceivable that crystal-packing restraints determine the conformations of these loops in the crystal. It cannot be discounted, however, that the loop conformations observed in the crystal also predominate in solution but are not observed due to insufficient nuclear-Overhauser-effect constraints. The N-terminal (Met1–Asn5) and C-terminal regions (residues 80–87) are random coil, as are loops I and II. The N-terminal residues pack against helices I and II and the C-terminal tail through several hydrogen bonds: Leu3(N) … Tyr10(O); Leu3(O) … His40(NE2); Ser8(N) … Gly83(O) and Tyr10(O) … Lys85(N). In addition, a hydrogen bond between Asn26(ND2) … Tyr55(OH) connects the C-terminal end of helix I with helix III. Helix I contains two internal salt bridges that help stabilize the helical structure in addition to the hydrogen-bonding pattern: Lys20–Glu23 and Glu21–Lys24. There is also a salt bridge which links helix I to helix IV (Lys73–Glu12). The C-terminal tail packs against helix IV and the N-terminal region through hydrogen bonds, including Pro82(O) … Arg76(NE) as well as the interactions described above for the N-terminal region.

Table 2 Positions and lengths of α-helices in Im7

α-Helix	Residues	Number of residues
I	Glu12–Lys24	13
II	Asp32–Thr45	14
III	Thr51–Tyr56	6
IV	Glu66–Gln79	14

Comparison of Im7 structures

The following discussion focuses on comparing the structures of Im7 from this work with those published previously by Chak et al. [27], which was solved by multiple isomorphous replacement using two derivatives. The first two residues were reported to be disordered by Chak et al. [27]; in our case, we observe density for the main-chain atoms only. Any attempts to model side-chain density resulted in unacceptably high temperature factors in subsequent refinement. We observed no density for the C-terminal glycine residue, which was modelled in the Chak structure. Chak et al. [27] reported weak density for the side-chains of residues Asn5, Glu39, Lys43 and Asp59, and our maps are largely consistent with this and we confirm that the latter

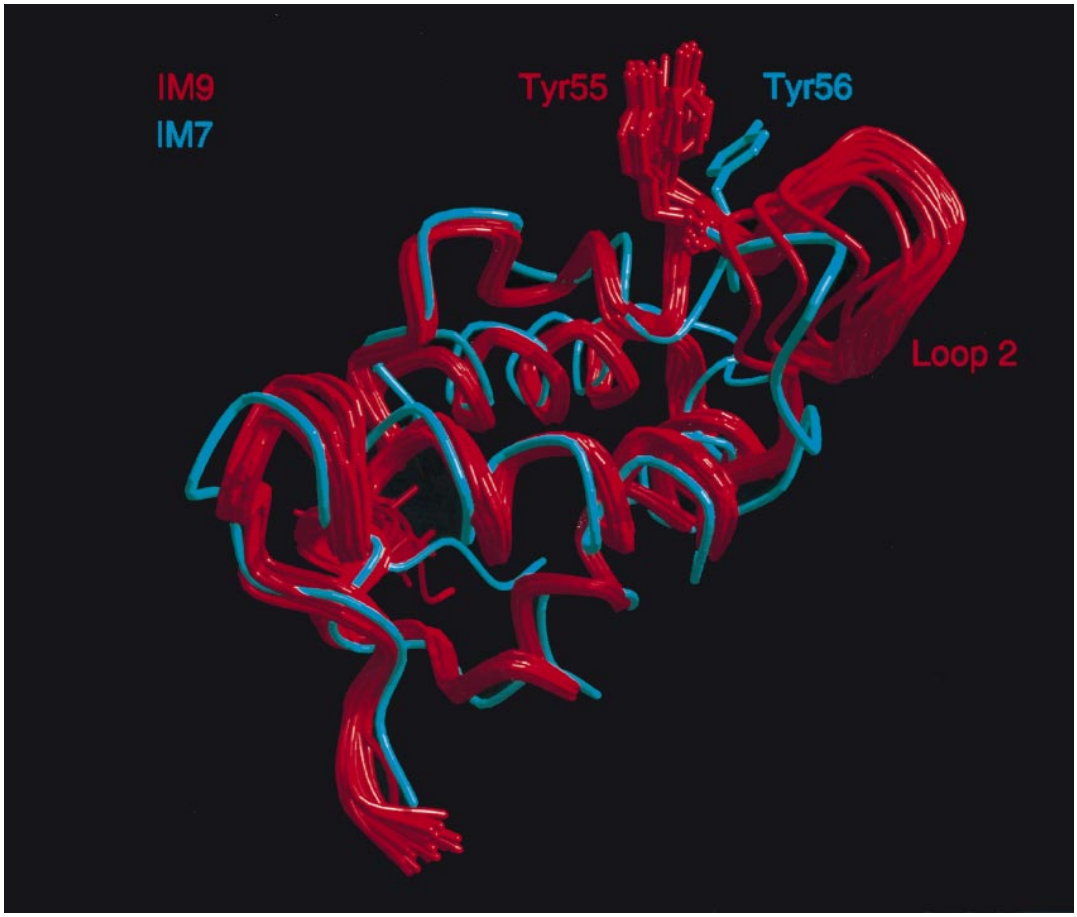

Figure 4 A conserved tyrosine that is essential for high-affinity E9 DNase binding is displayed differently from the Im7 and Im9 scaffolds

Shows the lowest energy structures for Im9, which satisfy the 1266 distance restraints used by Osborne et al. [24] to solve its solution structure, overlayed with the single crystal structure of Im7. This highlights the different orientation adopted by Tyr56 in Im7 compared with that in Im9 (Tyr55).

three residues are indeed mobile in this crystal form. We do not observe side-chain density for Ser58 and Lys81, and these side-chains have been modelled as alanines. We do, however, see clear density for Asn5. Chak et al. [27] report high average temperature factors (29.9 Å^2) for non-hydrogen atoms for their final model and very high average temperature factors for the loop regions (loop I, 38.4 Å^2; loop II, 43.5 Å^2), which is not the case here. In the present structure, Lys85 hydrogen bonds to Tyr10, not Asp9 as reported by Chak et al. [27], whereas Arg76 adopts a different conformation and does not hydrogen bond to Thr45(O) and Glu46(O), but rather to the C-terminal region. In summary, the structure of Im7 reported here is very similar to that determined by Chak et al. [27], thereby validating both structure determinations, one obtained through multiple isomorphous replacement and one through molecular replacement.

Comparing different immunity-protein structures

The sequences for Im7 and Im9 differ in length by one residue (Figure 1); loop I in Im7, linking helix I to helix II, contains an insertion of one residue relative to Im9. When compared with the NMR structure of Im9, the Im7 crystal structure determined here shares the same overall fold, and the two structures can be superimposed by least-squares alignment of equivalent Cα atoms with a global root-mean-squares deviation of 4.0 Å (Figure 3), which reduces to 1.7 Å if loops and termini are excluded. In the crystal structure, the N- and C-terminal regions pack against helices I and IV, which is not seen in the NMR structure.

The following analysis compares each element of secondary structure between the two immunity proteins. In helix I, side-chains from both structures adopt very similar conformations, despite some sequence differences. Residues contributing to the hydrophobic core show the greatest similarity. In helix II, there are some differences in the region 34–41; many of these residues differ in sequence and conformation, and this is likely to be important for the specificity differences of the immunity-protein family. Residue 35 (34 in Im9; Figure 1) has already been shown by mutagenesis to be a prime specificity determinant [50].

The majority of the residues in helix III, the shortest and most conserved element of secondary structure, superimpose in Im9 and Im7, except for a single tyrosine residue. The first tyrosine in helix III of Im7 (Tyr55; Tyr 54 in Im9) is in an equivalent position in the two proteins, except that in Im7 it forms a hydrogen bond with the asparagine residue at the base of helix I. At present, it is unclear from the NMR structure of Im9 if there is an equivalent hydrogen bond. The second tyrosine residue (Tyr56 in Im7, Tyr 55 in Im9) clearly adopts different conformations in the two structures (Figure 4). In Im9, it extends into the solvent, whereas in Im7 it folds back on to the helix. This difference is likely to be significant in the determination of

DNase specificity (see below). This residue forms a hydrogen bond with a neighbouring molecule in the crystal, but this is unlikely to be the reason for the observed difference in its orientation, for two reasons. First, it would appear from the structural comparison of Im7 with Im9 that the disposition of this residue is much affected by the conformation of the adjoining loop 2 (Figure 4). Loop 2 itself is not involved in binding the DNase [25,51], but substituting loop 2 between immunity proteins is known to affect colicin specificity, albeit only marginally [26]. This is consistent with an indirect effect on binding, most likely through changing the orientation of Tyr56. Secondly, rotation of Tyr56 of Im7 about the Cα–Cβ bond does not bring it into the close proximity of the equivalent side chain in Im9.

In helix IV, equivalent side chains have the same orientation with one exception, the tryptophan residue situated towards the C-terminal end of the helix. The orientation of this conserved tryptophan residue in Im7 and Im9 differs, but this residue is not thought to play a role in DNase specificity.

Electrostatics

Colicin endonucleases are very basic proteins possessing pI values of ≈ 10.5, whereas the immunity proteins are acidic in nature with pIs of ≈ 4.3. Stopped-flow kinetic analysis of Im9 binding the E9 DNase has shown that their association follows a two-step mechanism in which an initial encounter complex further rearranges to form the final stable complex [16]. The association rate constant is diffusion controlled ($k_{on} = 4 \times 10^9$ $M^{-1} \cdot s^{-1}$) and highly salt-dependent [16], implying that the two proteins are steered electrostatically towards one another, as has been proposed for other complexes, such as that between the RNase barnase and its inhibitor, barstar [52].

All these features are consistent with electrostatics playing a role in colicin–immunity-protein interactions. It is of interest then to analyse the surface potentials of immunity proteins to address this further. In Im9, the area of most concentrated negative charge is centred around helices II and III (Figure 5a), which coincides with the E9 DNase binding site [25,51]. On analysing the surface potentials of Im7, however, it can be seen that the most negatively charged areas, composed of acidic residues from the C-terminal end of helix I, loop I, the N-terminal region of helix II and loop II, do not correspond precisely with those of Im9 (Figures 5c and 5d). Most of these residues are variable in sequence in the immunity-protein family and this, together with the degree of negative charge at these sites, led Chak et al. [27] to propose these regions as the DNase binding site. However, homologue-scanning mutagenesis between Im2 and Im9 demonstrates unambiguously that neither helix I nor any of the loops plays a direct role in specificity [26]. It is possible that Im9 and Im7 bind the DNase with different surfaces, but this seems unlikely considering the model of immunity binding that is emerging; all the immunity proteins can bind and inhibit the E9 DNase, so it is more likely that they do so by similar mechanisms. It follows then that similar regions of the protein are involved and, from work on Im9 [51], these are likely to encompass helices II and III of each immunity protein, regardless of the overall distribution of negative surface potential. An interesting consequence of this, in terms of DNase binding, is that non-cognate immunity proteins such as Im2 and Im8, which have distributions of negative charges different to those of Im9, display very similar E9 DNase association kinetics [19]. Thus, the overall distribution of negative and positive charges in these proteins does not seem to dramatically influence their rate of association. This distribution does, however, affect immunity specificity, but only for those residues displayed from helices II and III.

Immunity specificity

The equilibrium dissociation constant for Im7 binding the E9 DNase is 10^{-4} M, which is the weakest of all the non-cognate immunity proteins, and indeed this does not provide over-expressing cells with any biological protection towards the action of ColE9 [19]. Im9 binds the same toxin with a K_d of $< 10^{-16}$ M, hence the specificity difference with respect to Im7 corresponds to differences in dissociation constants for the E9 DNase of 12 orders of magnitude. The question addressed in the present study is, do the structures of the isolated proteins provide a rationale for this extraordinarily wide range of binding affinities? In conjunction with some recent mutagenesis data, the structural comparison presented here suggests some strong possibilities.

We have completed an alanine scan across the putative E9 DNase binding site of Im9 [51], first identified by ^{15}N-isotope-edited NMR experiments as encompassing helices II and III [25]. This work showed that five residues of helix III contribute over two thirds of the DNase binding energy, while a further four to five residues of helix II contribute the remaining one third. Moreover, this work showed that three residues in helix III (Asp51, Tyr54 and Tyr55) each perturb binding by > 5 kcal/mol when substituted for alanine. These residues are conserved in this family of proteins whereas those of helix II, which governs specificity [26], are variable in sequence. We have interpreted the pattern of binding energies across these helices in terms of a 'dual recognition' mechanism of specificity, in which highly stabilizing anchor interactions emanating from the conserved protein scaffold of an immunity protein are modulated by the variable residues of helix II. Considering the high degree of sequence identity between the immunity proteins, the 'dual recognition' hypothesis proposed originally that residues from the conserved scaffold of helix III would be presented similarly for each immunity protein. The structural comparison of Im7 with Im9 shows that this is certainly true for Asp52 (Asp51 in Im9) and Tyr55 (Tyr54 in Im9). However, this is not the case for Tyr56 (Tyr55 in Im9), which occupies a different position in the two proteins (Figure 4). Differences such as this may explain the weak binding of Im7 to the E9 DNase; either this residue does not dock correctly on the enzyme or there are significant entropic penalties when the side-chain is reoriented so that it can dock appropriately.

The conserved residues of helix III, which make up the E9 DNase binding site, lie close to the specificity residues of helix II and we have proposed that interactions from these conserved sites are modulated by the residues of helix II [51]. This modulation could come about by the insertion of inappropriate helix II residues into binding pockets on the enzyme. For example, Val34 of Im9 is an aspartic acid in Im7 (Asp35), and Glu41 in Im9 is a valine in Im7 (Val42). Substitution of these residues in Im9 for alanine has a significant effect on binding the E9 DNase; the Val34 $\rightarrow$ Ala mutation decreases binding by > 3 kcal/mol, whereas the Glu41 $\rightarrow$ Ala mutation decreases binding by ≈ 2 kcal/mol. In the Im9–E9 DNase complex, it is likely that Val34 of Im9 docks into a well-packed hydrophobic recognition site on the enzyme, since its substitution for alanine is of the same order as that expected for the removal of two methyl groups from the hydrophobic core of a protein [52]. Conversely, the docking site for the acidic side chain of Glu41 is expected to be polar. In the event that the conserved residues of helix III of a non-cognate immunity protein such as Im7 dock on the enzyme, the distribution of charged and hydrophobic residues along helix

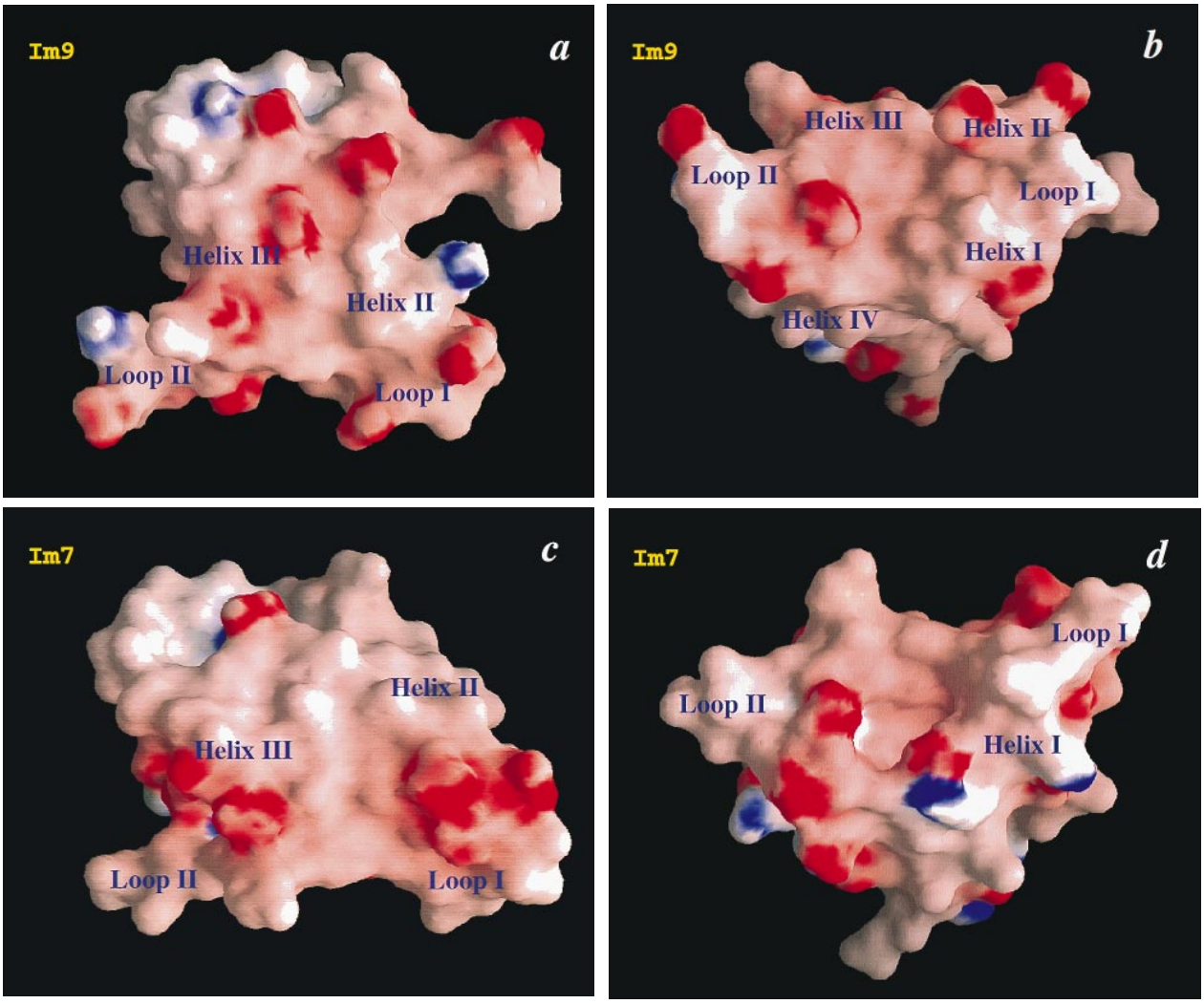

Figure 5 Electrostatic potentials of Im9 and Im7

Poisson-Boltzmann electrostatic potentials were calculated on a grid and interpolated on to the molecular surface using the program GRASP; blue-coloured regions indicate areas of the surface of the molecule with positive charge, whereas red indicates areas of negative charge. The orientation of the structures in (**a**) and (**c**), in which the DNase-binding surfaces of helices II and III are emphasized, is equivalent to that shown in Figure 3. (**b**) and (**d**) are views of (**a**) and (**c**) seen from underneath, i.e. an approximate 90° rotation around the horizontal axis. Note the different position of one of the tyrosines at the base of helix III in (**a**) and (**c**), as well as the different distribution of hydrophobic and acidic patches along helix II in the two immunity proteins.

II would not match the docking sites on the enzyme. Thus, Asp35 would be inserted into a hydrophobic cavity, whereas Val42 would inserted into a charged or polar environment, both interactions destabilizing the complex.

Conclusions

The crystal structure for Im7 has been solved by molecular replacement using an ensemble of conformers that describe the solution structure of Im9. Both the structure and the solution experiments indicate that Im7 is a monomer. A structural comparison between Im7 and Im9 reveals that a conserved tyrosine residue (Tyr56), which is energetically important for binding the cognate Im9–E9 DNase complex, is presented differently and this may be part of the mechanism of specificity. Modulation of the binding interactions emanating from the conserved residues of an immunity protein by residues from helix II is given further weight by this structural analysis. It shows that the distribution of hydrophobic and polar groups which juxtapose the anchor interactions of helix III in Im9 and Im7 are very different.

We thank Mike Osborne for the gift of the Im9 co-ordinates, Ann Reilly for expert technical assistance, Alec Tucker for guidance during the crystallization trials and Joe Jaeger for assistance with Figure 4. This work was funded by the U.K. Biotechnology and Biological Sciences Research Council. We thank the University of Oxford analytical ultracentrifugation facility for access to equipment.

REFERENCES

1 James, R., Kleanthous, C. and Moore, G. R. (1996) Microbiology **142**, 1569–1580
2 Ohno-Iwashita, Y. and Imahori, K. (1980) Biochemistry **19**, 652–659
3 Di Masi, D. R., White, D. C., Schnaitman, C. A. and Bradbeer, C. (1973) J. Bacteriol. **115**, 506–513
4 Pattus, F., Massotte, D., Wilmsen, H. U., Lakey, J., Tsernoglou, D., Tucker, A. and Parker, M. W. (1990) Experimentia **46**, 180–192

5 Stroud, B. (1995) Curr. Opin. Struct. Biol. **5**, 514–520
6 Cramer, W. A., Heyman, J. B., Schendal, S. L., Deriy, B. N., Choen, F. S., Elkins, P. A. and Stauffacher, C. V. (1995) Annu. Rev. Biophys. Biomol. Struct. **24**, 611–641
7 Gouaux, E. (1997) Structure **5**, 313–317
8 Boon, T. (1971) Proc. Natl. Acad. Sci. U.S.A. **68**, 2421–2425
9 Senior, B. W. and Holland, I. B. (1971) Proc. Natl. Acad. Sci. U.S.A. **68**, 959–963
10 Schaller, K. and Nomura, M. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 3989–3993
11 Jakes, K. S. and Zinder, N. D. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 3380–3384
12 Watson, R., Rowsome, W., Tsao, J. and Visentin, L. P. (1981) J. Bacteriol. **147**, 569–577
13 Cooper, P. C. and James, R. (1984) J. Gen. Microbiol. **130**, 209–215
14 Sidikaro, J. and Nomura, M. (1974) J. Biol. Chem. **249**, 445–453
15 Wallis, R., Reilly, A., Rowe, A., Moore, G. R., James, R. and Kleanthous, C. (1992) Eur. J. Biochem **207**, 687–695
16 Wallis, R., Moore, G. R., James, R. and Kleanthous, C. (1995) Biochemistry **34**, 13743–13750
17 Pugsley, A. P. and Schwartz, M. (1983) Mol. Gen. Genet. **190**, 366–372
18 Lau, C. K., Parsons, M. and Uchimura, T. (1992) in Bacteriocins, Microcins and Lantibiotics (James, R., Lazdunski, C. and Pattus, F., eds.), pp. 353–378, NATO ASI Series H, Springer, Heidelberg
19 Wallis, R., Leung, K.-Y., Pommer, A. J., Videler, H., Moore, G. R., James, R. and Kleanthous, C. (1995) Biochemistry **34**, 13751–13759
20 Wallis, R., Reilly, A., Barnes, K., Abell, C., Campbell, D. G., Moore, G. R., James, R. and Kleanthous, C. (1994) Eur. J. Biochem. **220**, 447–454
21 Garinot-Schneider, C., Pommer, A. J., Moore, G. R., Kleanthous, C. and James, R. (1996) J. Mol. Biol. **260**, 731–742
22 Curtis, M. D. and James, R. (1991) Mol. Microbiol. **5**, 2727–2733
23 Osborne, M. J., Lian, L.-Y., Wallis, R., Reilly, A., James, R., Kleanthous, C. and Moore, G. R. (1994) Biochemistry **33**, 12347–12355
24 Osborne, M. J., Breeze, A., Lian, L.-Y., Reilly, A., James, R., Kleanthous, C. and Moore, G. R. (1996) Biochemistry **35**, 9505–9512
25 Osborne, M. J., Wallis, R., Leung, K.-Y., Williams, G., Lian, L.-Y., Kleanthous, C. and Moore, G. R. (1997) Biochem. J. **323**, 823–831
26 Li, W., Dennis, C. A., James, R., Moore, G. R. and Kleanthous, C. (1997) J. Biol. Chem. **272**, 22253–22258
27 Chak, K.-F., Safo, M. K., Ku, W.-Y., Hsieh, S.-Y. and Yuan, H. S. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 6437–6442
28 Hsieh, S.-Y., Ko, T.-P., Tseng, M.-Y., Ku, W.-Y., Chak, K.-F. and Yuan, H. S. (1997) EMBO J. **16**, 1444–1454

Received 9 March 1998/17 April 1998; accepted 27 April 1998

29 Cohn, E. J. and Edsall, J. T. (1943) in Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, pp. 370–381, Reinhold, New York
30 Johnson, M. L., Correia, J. J., Yphantis, D. A. and Halvorson, H. R. (1981) Biophys. J. **36**, 575–588
31 Kabsch, W. (1988) J. Appl. Crystallogr. **21**, 67–71
32 Kabsch, W. (1993) J. Appl. Crystallogr. **26**, 795–800
33 Collaborative computational project, no. 4 (1994), Acta Cryst. **D50**, 760–763
34 Navaza, J. (1994) Acta Crystallogr. **A50**, 157–163
35 Navaza, J. and Saludjian, P. (1997) Methods in Enzymol. **276**, 581–594
36 Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. (1991) Acta Crystallogr. **A47**, 110–119
37 Murshodov, G. N., Dodson, E. J. and Vagin, A. A. (1996) in Macromolecular refinement: Proceedings of the CCP4 study weekend (Dodson, E., Moore, M., Ralph, A. and Bailey, S., eds.), pp. 93–104, Daresbury Laboratory, Warrington
38 Murshodov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Acta Crystallogr. **D53**, 240–255
39 Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993) J. Appl. Crystallogr. **26**, 283–292
40 Ku, W.-Y., Wang, C.-S., Chen, C.-Y., Chak, K.-F., Safo, M. K. and Yuan, H. S. (1995) Proteins Struct. Funct. Genet. **23**, 588–590
41 Crosio, M. P., Janin, J. and Jullien, M. (1992) J. Mol. Biol. **228**, 243–251
42 Janin, J. and Rodier, F. (1995) Proteins Struct. Funct. Genet. **23**, 580–587
43 Brünger, A. T., Campbell, R. L., Clore, G. M., Gronenborn, A. M., Karplus, M., Petsko, G. A. and Teeter, M. M. (1987) Science **235**, 1049–1053
44 Braun, W., Epp, O., Wuthrich, K. and Huber, R. (1989) J. Mol. Biol. **206**, 669–676
45 Baldwin, E. T., Weber, I. T., St Charles, R., Xuan, J. C., Appella, E., Yamada, M., Matsushima, K., Edwards, B. F. P., Clore, G. M., Gronenborn, A. M. and Wlodawer, R. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 502–506
46 Müller, T., Oehlenschlayer, F. and Buehner, M. (1995) J. Mol. Biol. **247**, 360–372
47 Weiss, M. S., Anderson, D. H., Raffioni, S., Bradshaw, R. A., Ortenzi, C., Luponni, P. and Eisenberg, D. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 10172–10176
48 Wilmanns, M. and Nilges, M. (1996) Acta Crystallogr. **D52**, 973–982
49 Thunnissen, M. M. G. M., Taddei, N., Liguri, G., Ramponi, G. and Nordlund, P. (1997) Structure **5**, 69–79
50 Wallis, R., Moore, G. R., Kleanthous, C. and James, R. (1992) Eur. J. Biochem. **210**, 923–930
51 Wallis, R., Leung, K.-Y., Osborne, M. J., James, R., Moore, G. R. and Kleanthous, C. (1998) Biochemistry **37**, 476–485
52 Ferscht, A. R. and Serrano, L. (1993) Curr. Opin. Struct. Biol. **3**, 75–83
Schreiber, G. and Fersht, A. R. (1993) Biochemistry **32**, 5145–5150