Three-Dimensional Solution Structure and ^{13}C Nuclear Magnetic Resonance Assignments of the Colicin E9 Immunity Protein Im9[†,‡]

Michael J. Osborne,[§] Alexander L. Breeze,[||] Lu-Yun Lian,[⊥] Ann Reilly,[▽] Richard James,[▽] Colin Kleanthous,[▽] and Geoffrey R. Moore*[,§]

Schools of Biological and Chemical Sciences, University of East Anglia, Norwich NR4 7TJ, U.K., Protein Structure Group, Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire SK10 4TG, U.K., and Biological NMR Centre, Leicester University, Leicester, U.K.

Received February 20, 1996; Revised Manuscript Received May 13, 1996[⊗]

ABSTRACT: The 86-amino acid colicin E9 immunity protein (Im9), which inhibits the DNase activity of colicin E9, has been overexpressed in *Escherichia coli* and isotopically enriched with ^{15}N and ^{13}C. Using the 3D CBCANH and CBCA(CO)NH experiments, we have almost completely assigned the backbone ^{13}C resonances and extended previously reported $^{15}N/^{1}H$ backbone assignments [Osborne *et al*. (1994), *Biochemistry 33*, 12347−12355]. Side chain assignments for almost all residues were made using the 3D ^{13}C HCCH-TOCSY experiment allied to previous ^{1}H assignments. Sixty solution structures of Im9 were determined using the DIANA program on the basis of 1210 distance restraints and 56 dihedral angle restraints. The 30 lowest-energy structures were then subjected to a slow-cooling simulated annealing protocol using XPLOR and the 21 lowest-energy structures, satisfying the geometric restraints chosen for further analysis. The Im9 structure is well-defined except for the termini and two solvent-exposed loops between residues 28−32 and 57−64. The average RMSD about the average structure of residues 4−84 was 0.94 Å for all heavy atoms and 0.53 Å for backbone C^{α}, C═O, and N atoms. The Im9 fold is novel and can be considered a distorted antiparallel four-helix bundle, in which the third helix is rather short, being terminated close to its N-terminal end by a proline at its C-terminus. The structure fits in well with available kinetic and biochemical data concerning the interaction between Im9 and its target DNase. Important residues of Im9 that govern specificity are located on the molecular surface in a region rich in negatively charged groups, consistent with the proposed electrostatically steered association [Wallis *et al.* (1995a), *Biochemistry 34*, 13743−13750].

Colicin E9 (Col E9) is a plasmid-encoded DNase that is secreted as part of the stress response system of some *Escherichia coli* (Luria & Suit, 1987; Eaton & James, 1989). It binds to the vitamin B_{12} extracellular receptor encoded by the *btuB* gene on target cells (Di Masi *et al.*, 1973), and, following translocation across the outer and cytoplasmic membranes, its cytotoxic activity leads to cell death. To provide the host organism with immunity from the action of colicin E9, an inhibitor protein, Im9, is coordinately produced (Chak & James, 1986; James *et al.*, 1987). This acidic protein binds to intact Col E9 and its DNase domain with a dissociation constant of $\sim 10^{-16}$ M, revealing the complex to be one of the tightest interprotein complexes known (Wallis *et al.*, 1995a). It is the complex of Col E9 with Im9 that is released by the producing cell, but the complex is presumed to be dissociated during its interaction with the target cell since the DNase is active within it. In addition to Col E9, colicin DNases E2, E7, and E8 have been isolated, all with their own specific inhibitor proteins: Im2, Im7, and Im8 respectively (Schaller & Nomura, 1976; Toba *et al*., 1988; Chak *et al*., 1991). There is a high degree of sequence similarity between the four colicin DNase domains (∼80%) and between their inhibitor proteins (∼50%), but despite this, only the cognate inhibitors provide complete immunity against the colicin DNase action (James *et al*., 1992; Wallis *et al.*, 1995b). However, noncognate complexes are formed between the E9 DNase domain and Im2, Im7, and Im8 with K_d values ranging from $10^{-4}-10^{-8}$M (Wallis *et al.*, 1995b).

The combination of high binding affinities between cognate partners and relatively low binding affinities between noncognate proteins, coupled with the relatively small size of the complex formed between a colicin DNase domain and an inhibitor protein (220 amino acids for Col E9/Im9), makes this an attractive system for investigating molecular recognition events between proteins. To pursue this, we are engaged in a detailed kinetic, protein engineering, and NMR study of the Col E9/Im9 proteins, both in their free states and in their interprotein complex. We have overexpressed the 134-amino acid DNase domain of Col E9 (Wallis *et al.*, 1994) and the 86-amino acid Im9 (Wallis *et al.*, 1992a) and have used NMR spectroscopy to obtain structural information concerning both free and bound Im9 (Osborne *et al.*, 1994). We have previously reported NMR studies on uniformaly ^{15}N-labeled Im9 that provided assignments for 77 of its 86

[†] This work was supported by the Wellcome Trust (a Research Leave Fellowship to G.R.M.) and by the Biotechnology & Biological Sciences Research Council.

[‡] Coordinates have been deposited in the Brookhaven Protein Data Bank for the 21 final structures (1IMP) and the average structure (1IMQ).

* Author to whom correspondence should be addressed. FAX: +44-1603 259396. E-mail: g.moore@uea.ac.uk.

[§] School of Chemical Sciences, University of East Anglia.

[||] Zeneca Pharmaceuticals.

[⊥] Leicester University.

[▽] School of Biological Sciences, University of East Anglia.

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1996.

S0006-2960(96)00401-1 CCC: $12.00

amino acid residues (Osborne *et al.*, 1994). That study also showed that the protein contained at least three helical sections, with the specificity-determining region forming one of the helices. However, the incomplete assignments, coupled with severe overlap in the aliphatic region of the $^1H-^1H$ NOESY spectrum, prevented determination of a reliable three-dimensional (3D) structure at that stage, and to overcome these problems, we embarked on a study of uniformly ^{13}C- and ^{15}N-labeled Im9. In the present paper, we report further ^1H and ^{15}N assignments of Im9, together with ^{13}C assignments, and describe its three-dimensional structure determined using NOE and backbone dihedral angle constraints obtained from multidimensional heteronuclear NMR data.

MATERIALS AND METHODS

Sample Preparation. Im9 was expressed in *E. coli* and purified as previously described (Wallis *et al*., 1992b). All samples prepared for NMR were assayed for biological activity as described by Wallis *et al*. (1992b) and found to be fully active. Uniformly ^{15}N-labeled Im9 was obtained by growing the *E. coli* JM105 cells containing the expression system in minimal medium with $^{15}NH_4Cl$ (1 g/L), and uniformly ^{13}C- and ^{15}N-labeled Im9 was obtained from cells grown in minimal medium containing a mixture of [$^{13}C_6$]glucose (4 g/L) and $^{15}NH_4Cl$ (1 g/L). Yields of protein from these enrichments were typically 20 mg/L. Concentrations of NMR samples in either 90% ^1H_2O/10% ^2H_2O or 99.9% ^2H_2O were typically 1−4 mM. The NMR solutions were prepared from Im9 lyophilized from either ^1H_2O or ^2H_2O solutions containing sodium phosphate at pH 7.0 or 6.2 to yield NMR samples with a final phosphate concentration of 25−50 mM.

NMR Measurements. The NMR data were acquired on Bruker AMX600 or Varian Unity 600 spectrometers. Spectra measured with the Bruker instrument were recorded in the pure phase absorption mode by the time-proportional phase incrementation (TPPI) method (Redfield & Kunz, 1975; Marion & Wüthrich, 1983), and on the Varian instrument, quadrature detection was achieved in the indirectly acquired dimensions by the States−TPPI procedure (Marion *et al.*, 1989b). Proton chemical shifts were measured from internal dioxane at 3.77 ppm; nitrogen chemical shifts were referenced using external $^{15}NH_4NO_3$ (5 M $^{15}NH_4NO_3$ in 2 M HNO_3) with the ammonium ^{15}N resonance taken as 0 ppm, and carbon chemical shifts were referenced to the methyl resonance of external sodium 3-(trimethylsilyl)-propionate. Experimental details of the homonuclear and $^1H-^{15}N$ experiments are given in Osborne *et al*. (1994). NOESY (Jeener *et al*., 1979: Macura *et al*., 1981) data were collected using mixing times of 100 and 150 ms. Additional details of such experiments carried out for the present paper and details of the triple-resonance experiments are given below.

All heteronuclear 3D experiments were carried out at 293 K. The $^1H-^{15}N$ 3D TOCSY-HMQC, NOESY-HMQC (Zuiderweg & Fesik, 1989; Marion *et al*., 1989a), and HMQC-NOESY-HMQC (Frenkiel *et al*., 1990) pulse schemes used were similar to those originally described. These were all measured on the Bruker instrument. The CBCANH and CBCA(CO)NH spectra were acquired with the Varian spectrometer using pulse schemes similar to those reported by Grzesiek and Bax (1992a,b, 1993). The data set sizes were 512 (^1H) × 56 (^{13}C) × 30 (^{15}N) complex points with 16 scans recorded for the CBCA(CO)NH spectrum and 48 scans for the CBCANH spectrum. Water suppression was achieved with a 2.5 ms spin lock purge pulse. The CN-NOESY-HSQC data were acquired with the pulse scheme of Pascal *et al.* (1994) on the Varian spectrometer, using pulsed-field gradients for water suppression and artifact reduction as described by Pascal *et al.* (1994). ^{13}C(carbonyl) decoupling during t_2 was accomplished using a cosine-modulated SEDUCE-1 decoupling waveform (McCoy & Mueller, 1992) applied off-resonance at the ^{13}C(aliphatic/aromatic) carrier, positioned at 67ppm. Acquisition times of 49 ms (t_3,^1H), 15.1 ms (t_1, ^1H), and 11.8 ms (t_2, $^{15}N/^{13}C$) were employed with mixing times of 100 and 150 ms. Data were processed on Silicon Graphics Indigo or Sun Sparc workstations using FELIX (Biosym/MSI). For the triple-resonance and CN-NOESY-HSQC data sets, conventional or "mirror image" (Zhu & Bax, 1990) linear prediction methods were employed to extend the data in the indirectly acquired dimensions where necessary.

Structural Restraints and Calculations. A total of 1307 unique NOE cross-peaks were assigned from the 3D CN-NOESY spectrum of a doubly labeled sample of Im9 recorded with a mixing time of 100 ms in ^1H_2O at 295 K. Peak volumes were translated into upper limits of interatomic distances using the CALIBA program and structures generated using DIANA, by the protocol described by Güntert *et al*. (1991). For the initial structure calculations, cross-peak volumes were calibrated from resonances arising from regions of the structure which were presumed to be relatively rigid (e.g. sequential amide proton connectivities at the center of α-helices). After the generation of initial structures using DIANA (Güntert *et al*., 1991), peak volumes were calibrated independently for the five calibration classes described by Güntert *et al*. (1991). To account for effects of spin diffusion and ^{13}C relaxation times, a maximum upper distance limit of 6 Å was used. One hundred thirty-five of the distance restraints were deemed meaningless (i.e. the constraint defined a fixed distance or would not violate any possible conformation), leaving 1172 distance constraints from NOE data. Fifty-six $^3J_{NH\alpha}$ coupling constants were measured from an HMQC-J experiment and converted to ϕ torsion restraints as follows. For $^3J_{NH\alpha}$ of >8 Hz, ϕ was set to $-120 \pm 40°$, and for $^3J_{NH\alpha}$ values of <6 Hz, ϕ was set to $-65 \pm 25°$. Hydrogen bond restraints, 38 in total, were included after initial structure calculations unambiguously defined hydrogen bond acceptors for amide protons exchanging slowly with deuterons (Osborne *et al*., 1994). A single distance restraint was included for each hydrogen bond, HN > O (2.3 ± 0.5 Å). No angle restraints were determined for β-methylene protons. Sixty structures were generated using the DIANA program. The 30 lowest-energy structures which had no common NOE-derived distance violations greater than 0.5 Å and no common torsion angle violations greater than 5° were then subjected to a slow-cooling simulated annealing protocol (initial temperature of 2000 K, 10 000 cooling steps) (Nilges *et al*., 1988) using the XPLOR program (Brunger, 1992). Twenty one of these structures satisfied all geometric restraints (no NOE >0.5 Å and no dihedral >5°), and from this ensemble, an average structure was computed and subjected to an additional 1000 cycles of Powell restrained energy minimization. Analysis of the structures was carried out with MOLMOL (Koradi *et al*., 1996) and QUANTA (Molecular Simulations).

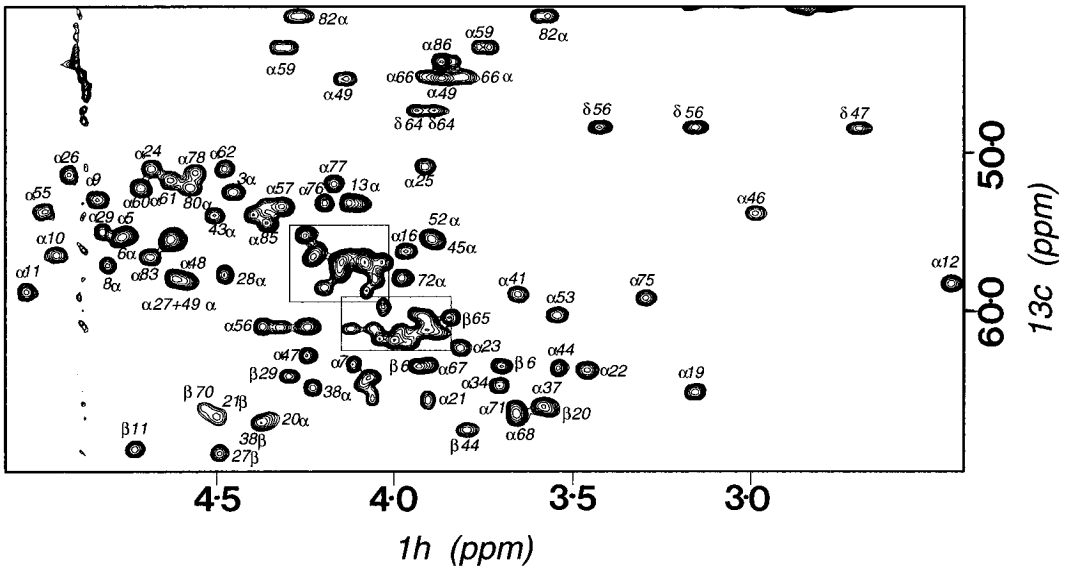

FIGURE 1: Selected region of the 2D $^{1}H-^{13}C$ HSQC spectrum of ^{13}C- and ^{15}N-labeled Im9. The boxed areas indicate regions where resonance overlap was a problem.

RESULTS AND DISCUSSION

Resolution of Resonances. Severe overlap of resonances in 2D $^{1}H-^{1}H$ spectra of Im9 necessitated the use of 3D ^{15}N-edited experiments to obtain sequential assignments following the procedure of Wüthrich (1986). Using this approach, ~93% of the backbone ^{1}H resonances of Im9 were assigned and the secondary structure delineated on the basis of short range and medium range NOEs. These showed at least three helices and no β-sheet to be present in Im9 (Osborne *et al*., 1994). Highly helical proteins tend to display poor backbone proton chemical shift dispersion (Wishart *et al*., 1992), and this was also observed for Im9. Chemical shift degeneracy of, in particular, α and β protons and methyl resonances meant that only poor resolution structures of Im9 were obtained from experiments utilizing ^{15}N-enriched samples of Im9. In general, overlap in the aliphatic region may be alleviated by editing spectra according to the ^{13}C chemical shifts of the directly bound carbon nuclei, and once assignments of ^{13}C chemical shifts are obtained, the often unique combination of ^{1}H and ^{13}C chemical shifts can be used to identify otherwise ambiguous NOEs in the ^{15}N- edited 3D NOESY spectra (see later). The resolution afforded by the introduction of the ^{13}C dimension for Im9 is evident from Figure 1 which shows a selected region of the $^{1}H-^{13}C$ HSQC spectrum of Im9. The majority of the 58 protons that resonate between 3.8 and 4.3 ppm in the ^{1}H NMR spectrum are resolved, though there remain some areas, indicated by boxes in Figure 1, where overlap is still a problem. Assignment of these areas was achieved using triple-resonance experiments, with a ^{13}C- and ^{15}N-enriched sample, which make use of the sequential and intraresidue ^{1}H and ^{15}N amide chemical shifts to separate the resonances. However, despite the assignments, ambiguity remains in assigning NOE cross-peaks from NOESY experiments, as described later.

Assignment of Resonances. Assignments of the backbone ^{13}C, ^{15}N, and ^{1}H resonances were obtained using a strategy similar to that of Grzesiek and Bax (1993). In this procedure, comparison of spectra from experiments which directly correlate peptide amide resonaces with the preceding interresidue $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances [CBCA(CO)NH experiment] (Grzesiek & Bax, 1992b) and with both the intraresidue $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances and the preceding interresidue $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances (CBCANH experiment) (Grzesiek & Bax, 1992a) affords sequential assignment of residues. Amino acid type is then identified on the basis of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts.

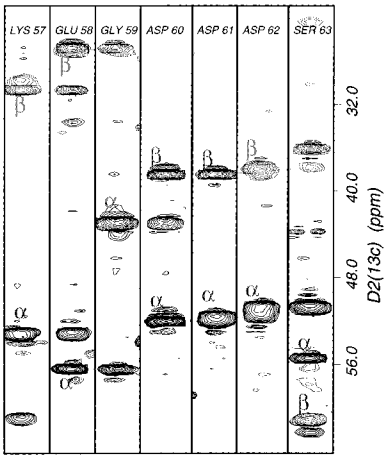

FIGURE 2: Region of a 600 MHz $^{13}C-^{15}N-^{1}H$ CBCANH spectrum of uniformly ^{15}N- and ^{13}C-labeled Im9 in 90% $^{1}H_2O$/10% $^{2}H_2O$ illustrating sequential assignments via heteronuclear scalar couplings. The figure shows 2D strips taken through the 3D CBCANH spectrum for residues Lys57−Ser63 in which the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of residue i are correlated with their intraresidue NH ^{15}N and ^{1}H chemical shifts and also the NH ^{15}N and ^{1}H chemical shifts of residue $i + 1$. Each strip represents a narrow D1(^{1}H) region of a D2(^{13}C)/D1(^{1}H) cross section of the 3D spectrum at the NH $^{1}H/^{15}N$ frequency of an amino acid. Intraresidue $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ connectivities are marked. $^{13}C^{\alpha}$ connectivities, except those of glycine, are positive (solid contours) and $^{13}C^{\beta}$ connectivities and glycine $^{13}C^{\alpha}$ connectivities are negative (broken contours).

Comparison of the CBCANH spectrum (Figure 2) with the CBCA(CO)NH spectrum (not shown) confirmed the previously reported ^{1}H and ^{15}N assignments (Osborne *et al*., 1994) and allowed almost complete assignments for the ^{1}H, ^{13}C, and ^{15}N backbone resonances of Im9 to be obtained. These assignments include those for the previously unassigned residues 62−64 (Figure 2) and 85−86 and the three proline residues, 47, 56, and 64. The $^{13}C^{\beta}$ shifts of the proline residues correspond to the random coil shifts for the trans configuration (Jardetzky & Roberts, 1981), which is in accordance with NOE analysis and the final solution

Table 1: Summary of NMR Parameters for Structure Calculations

restraint type	number of restraints
intraresidue NOEs	358
interresidue sequential NOEs ($\|i-j\|=1$)	311
interresidue short range NOEs ($1<\|i-j\|<5$)	269
interresidue long range NOEs ($\|i-j\|\geq 5$)	234
total number of NOEs	***1172***
hydrogen bond restraints (one per H bond)	38
ϕ dihedral angle restraints	56
total number of restraints	***1266***

structure. The information from the triple-resonance experiments was important, particularly the CBCANH experiment which displayed both inter- and intraresidue connectivities for residues 3−86.

The remaining side chain ^{13}C and ^{1}H assignments were identified from the 3D $^{13}C-^{1}H$ HCCH-TOCSY experiment in conjunction with previously assigned proton resonances (Osborne *et al*., 1994). This experiment afforded additional proton assignments, particularly for the glutamate and lysine side chains, and allowed a number of ambiguous assignments to be confirmed. The mixing time employed for this experiment, 20 ms, did not allow efficient carbon−carbon magnetization transfer between the C^{α} and C^{β} resonances of serine residues (Clore *et al*., 1990), and thus, unobservable, or weak, cross-peaks for these residues resulted. Additional serine proton resonances were assigned on the basis of their connectivities with $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances detected in the CN-NOESY spectra.

Assignment of NOEs. In our previous paper (Osborne *et al*., 1994), 3D ^{15}N-edited NOESY-HMQC and HMQC-NOESY-HMQC experiments were used to identify a large number of sequential and medium range NOEs involving amide and backbone protons and amide−amide protons. However, overlap in the aliphatic ^{1}H region still caused assignment problems. For example, in the α-helices, there are a large number of degenerate H^{α} and H^{β} protons, making it difficult to distinguish between sequential and medium range αN, βN, or αβ connectivities, which are important for defining helices (Wüthrich, 1986). More importantly, severe overlap often made it impossible to identify many long range NOE connectivities involving backbone atoms or methyl groups which also showed degeneracy in the ^{1}H dimension. Indeed, from ^{15}N-edited experiments, only 673 unambiguous NOE cross-peaks were identified, the majority of which were sequential or medium range and insufficient to define a high-resolution structure. We therefore utilized the ^{13}C dimension as well by collecting a 3D CN-NOESY-HSQC data set (Pascal *et al*., 1994). This experiment overcomes ^{1}H chemical shift degeneracy according to the directly bound ^{15}N or ^{13}C nucleus by simultaneously detecting ^{13}C and ^{15}N frequencies in one dimension. Thus, protons attached to degenerate ^{13}C or ^{15}N shifts are clearly resolved. This experiment yields similar information to the 4D ^{15}N- and ^{13}C-edited NOESY experiment (Kay *et al*., 1990), but since only three dimensions are present, it affords better digital resolution and thus sensitivity. It was particularly useful for identifying NOEs involving amide and backbone protons which were previously ambiguous and in identifying NOE connectivities involving methyl resonances. This experiment helped unambiguously assign a further 634 unique NOE cross-peaks, making a total of 1307 assigned NOE cross-peaks, of which 1172 were used for structure determination (Table 1). In regions of the spectrum where

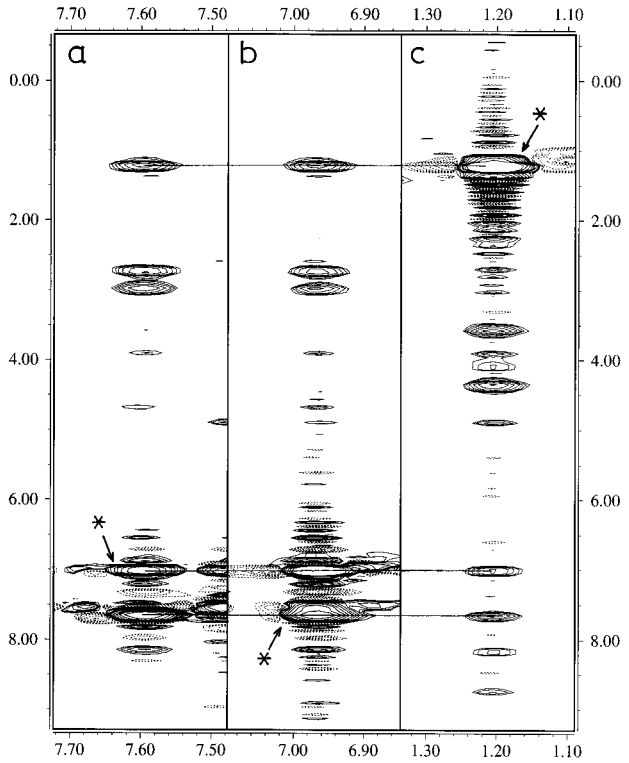

FIGURE 3: Selected 2D $^{1}H-^{1}H$ slices taken from a 150 ms 3D CN-NOESY-HSQC data set of ^{13}C- and ^{15}N-labeled Im9 in 90% $^{1}H_2O$/10% $^{2}H_2O$. The slices are at F_2 frequencies corresponding to the ^{15}N chemical shift of the side chain NH protons of Asn 24 (a and b) and the ^{13}C chemical shift of the γ-methyl protons of Thr20 (c). The diagonal cross-peaks are indicated by an asterisk, and the symmetry-related cross-peaks, showing NOEs between these protons, connected by a horizontal line.

overlap persists and was not resolved by the CN-NOESY experiment (for examples, see Figure 1), initial structure calculations were used to remove ambiguity. This involved structure determination using DIANA leaving out some of the NOESY cross-peak data and the derived structure used to assign unambiguous NOESY cross-peaks. The distance constraints table was then revised to include the new assignments, and structure calculations to assist assignment were again carried out. This procedure was repeated until no further ambiguous NOESY cross-peaks could be assigned. There remained NOESY cross-peaks where firm assignments could not be made even with the aid of the structure, and therefore, these NOEs were not used in the calculations. These included, for example, many NOE connectivities involving the α-carbons of residues His5 and Ser6 and the α-carbons of residues 30−33. An example of the excellent resolution afforded by the 3D CN-NOESY-HSQC experiment is shown in Figure 3 to illustrate a set of nonsequential NOE connectivities.

Chemical Shift Analysis. The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts can be used to identify amino acid type and are exquisitely sensitive to secondary structure in folded proteins. The chemical shift index (CSI) proposed by Wishart *et al*. (1992) and Wishart and Sykes (1994) can predict protein secondary structure on the basis of analyses of $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ chemical shifts relative to their random coil values with a predictive accuracy in excess of 92%. $^{13}C^{\alpha}$ resonances for residues in helices tend to be shifted >0.7 ppm to higher frequency than their random coil values. For β-sheets, perturbations of >0.7 ppm to lower frequency are usually seen for $^{13}C^{\alpha}$ peaks. $^{1}H^{\alpha}$ resonances in helices tend to exhibit

perturbations to lower frequency (>1.0 ppm) and vice versa for β-sheets. Continuous stretches of these perturbations are used to delineate secondary structure. The consensus CSI for Im9 is given in Figure 4a. This analysis indicates Im9 contains four α-helices involving residues 11−23, 28−42, 48−52, and 65−77. Such an analysis agrees well with the secondary structure of Im9 determined from the DIANA/XPLOR structure calculations (Figure 5).

Structural Restraints and Statistics. The structural statistics for the 21 converged structures (Figure 6) calculated from the structural restraints given in Table 1 are reported in Table 2. The final 21 structures exhibit an atomic RMSD about the average of 0.64 Å for the backbone C^{α}, C═O, and N atoms and 1.12 Å for all non-hydrogen atoms, indicating that the structures are reasonably well-defined by the experimental restraints. Additionally, deviations from idealized covalent geometry are very small, indicating that no severe distortions exist in the 21 structures. A more detailed assessment of local variability of atoms is provided by a plot of RMSD versus residue number (Figure 4c). This figure shows that the majority of the backbone atoms are well-defined, in particular the four helices whose backbone atoms exhibit an RMSD of 0.28 Å. Backbone atoms which are less well-defined are the termini and segments comprising residues 27−31 (which still exhibit RMSDs below 1 Å) and residues 56−62. These residues are relatively poorly defined by the experimental restraints (Figure 4d) and are also found at the surface of the protein (Figure 4b). Leaving out the terminal residues from the comparison of 21 structures indicates that the core structures are relatively well-defined: atomic RMSDs for residues 4−84 about the average are 0.53 Å for the backbone C^{α}, C═O, and N atoms and 0.94 Å for all non-hydrogen atoms. The local RMSD for the side chains is also well-defined, with ∼70% showing RMSD values below 1 Å. Residues whose side chains exhibit RMSDs above 1 Å are mainly located at the surface and are therefore likely to exhibit multiple conformations. The lack of NOE data at the termini and in the two loop regions may reflect fluctuations in these areas, and this point is addressed later.

The Ramachandran plot (Figure 7) for the restrained energy-minimized average structure further indicates the quality of the Im9 structure, with the majority of residues falling in the allowed regions of the plot. Residues 3, 4, 55, 60, 80, and 84 are outside the allowed regions. No residue falls into a disallowed region in all 21 experimental structures, and with the exception of residue 55, the residues with unfavorable dihedral angles in the average structure are located in loop regions or close to the termini and thus have correspondingly high RMSD values. Similar behavior was observed for some of the loop and terminal residues in the solution structure of barstar (Lubienski *et al*., 1994). Residues exhibiting positive ϕ values (25, 29, 45, 49, 59, 79, and 82) also appear in loop regions. This has been noted in other proteins and may reflect how energetically unfavorable ϕ angles allow short loops to connect secondary structural elements (Ludvigsen & Poulsen, 1992; Serrano *et al*., 1992; Lubienski *et al*., 1994).

Description of the Polypeptide Fold of Im9. The structure of Im9 (Figures 5 and 6) consists of three well-defined α-helices—helix I (Glu12−Cys23), helix II (Glu30−Thr44), and helix IV (Ser65−A77)—with an additional one-turn 3_{10} helix (Asp51−Tyr55). Helices I and II are antiparallel (forming an angle of −150°) and are connected by a loop comprising six residues which is relatively ill-defined. This

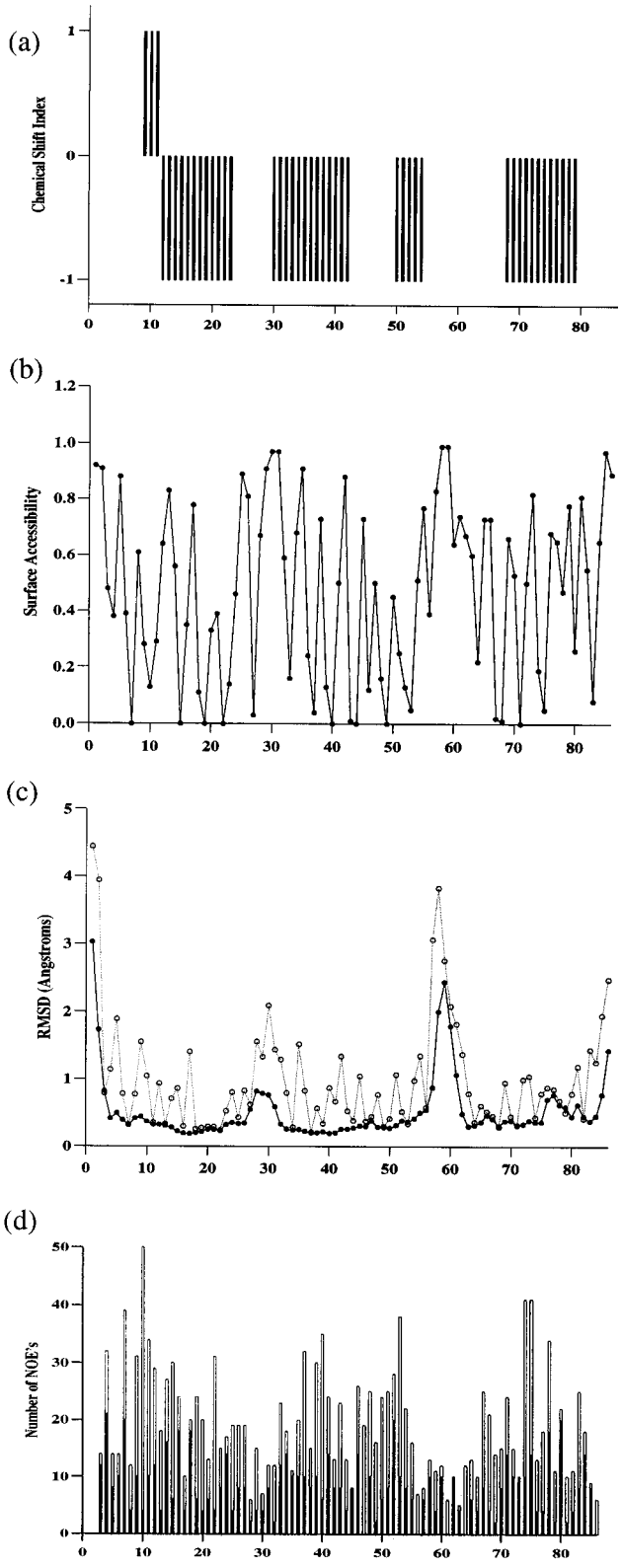

FIGURE 4: (a) Histograms showing the consensus chemical shift index of Im9 on the basis of the ^{13}C and ^{1}H chemical shifts (Wishart *et al*., 1992; Wishart & Sykes, 1994). (b) Relative surface accessibility per residue for a sphere with a 1.4 Å radius. (c) Distribution of the atomic RMSDs about the mean structure for the 21 solution structures of Im9 as a function of residue number for the C^{α}, C═O, and N backbone atoms (solid) and all non-hydrogen atoms (open). (d) Number of interresidue (shaded) and intraresidue (solid) NOEs per amino acid residue.

turn is highly solvent accessible as indicated by Figure 4b and NH exchange data; exchange throughout this stretch occurs within an 8 min period prior to recording of the first data set (Osborne *et al*., 1994). Amide exchange rates in

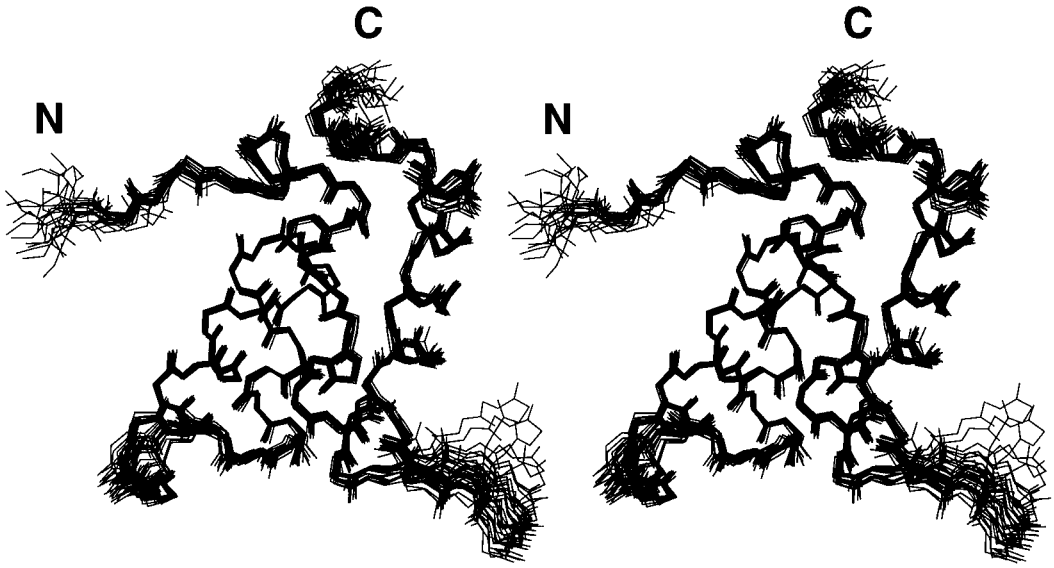

FIGURE 5: Stereoview showing the superposition of the C^{α}, C═O, and N backbone atoms for the 21 solution structures of Im9.

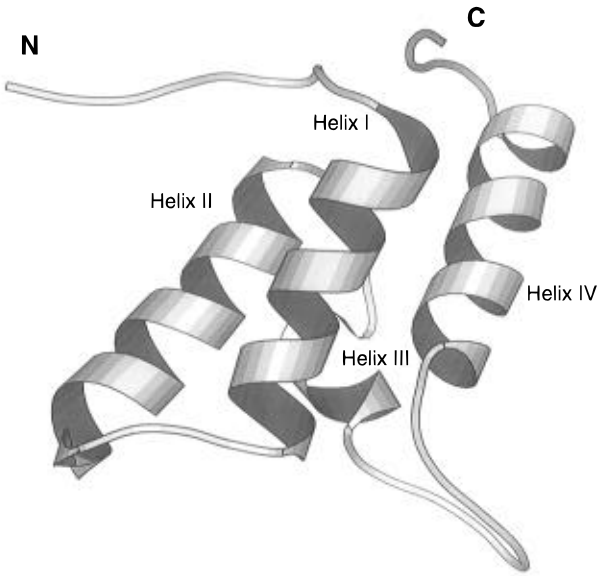

FIGURE 6: Ribbon structure of the energy minimized average structure of Im9 generated using the program MOLSCRIPT (Kraulis, 1991).

Table 2: Summary of Structure Statistics

RMS Deviation from Experimental Restraints	
distance	0.03 ± 0.001 Å
dihedral angle	$0.43 \pm 0.20°$
Deviations from Ideal Covalent Geometry Based on the XPLOR Force Field	
bonds	$4.97 \times 10^{-3} \pm 0.237 \times 10^{-3}$ Å
angles	$0.65 \pm 0.02°$
impropers	$0.44 \pm 0.02°$
Atomic RMS Differences from Average Structure	
backbone atoms	0.64 Å
all heavy atoms	1.12 Å

helix I tend to be the longest of the whole molecule (Osborne *et al*., 1994) and may reflect stabilization caused by a large number of contacts with helices II and IV and, to a lesser extent, helix III.

Helix II constitutes the longest helix in Im9. It is well-defined by the NMR data and forms contacts with the three other helices, but primarily with helix I. A well-defined β-turn between residues 45 and 49 connects helices II and III at an angle of 100° to each other. Again, positive ϕ angles

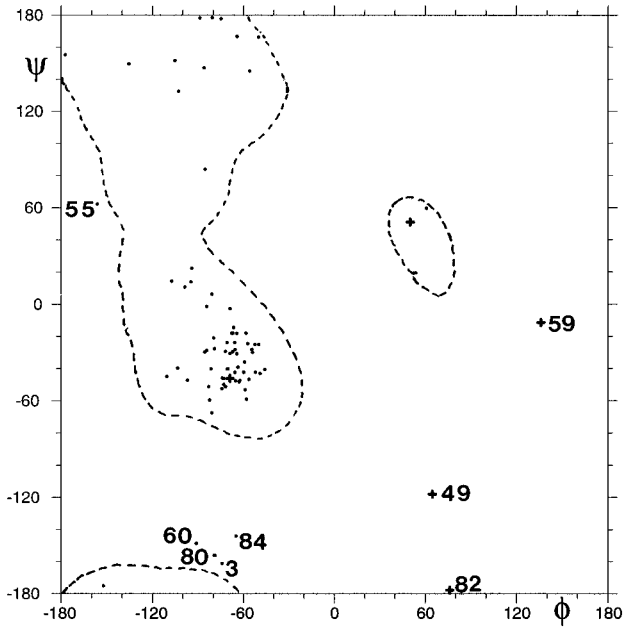

FIGURE 7: Ramachandran plot for the energy-minimized average structure of Im9 generated by MOLMOL (Koradi *et al*., 1996). The broken lines indicate the 95% allowed regions; + indicates glycine residue; and closed circles indicate other residues. Leu3, Tyr55, Asp60, Lys80, and Lys84 fall outside the 95% allowed regions (see text).

are observed for members of this connecting unit (Glu45 and Gly49). Despite its small size, helix III is well-defined by the NMR data and forms many contacts with the other helices. Additionally, the amide protons exchange relatively slowly (Osborne *et al*., 1994), although one might have expected them to be relatively rapidly exchanging since it is only a one-turn helix. The presence of bulky hydrophobic groups at the surface of this helix might protect the NH protons from the solvent.

Helices III and IV are connected by a long ill-defined loop incorporating residues 57−64, four of which have carboxylate side chains. Amide exchange data (Osborne *et al*., 1994) and surface accessibility calculations of the minimized average structure (Figure 4b) show that this loop, and many of its side chain atoms, are accessible to solvent. The poor definition in this area reflects a lack of restraining interresidue interactions in keeping with this loop being relatively

unconstrained in solution. This is consistent with the CSI analysis which indicates residues in this loop are in random configurations (Figure 4a). The final helix brings both termini close together to allow an H bond between the NH of Lys84 and the backbone O of Tyr10 which is reflected in amide exchange data. In addition to contacts with the other helices, helix IV has contacts with the loop linking helices II and III, with Trp74 particularly close to His46.

Side Chains. About 70% of the side chains are well-defined. Continuous stretches with RMSDs of 1.2 Å or greater are observed for the N- and C-termini and the loop regions involving residues 28−32 and 57−64. The side chains of these residues are situated at, or close to, the surface of Im9 and are therefore highly solvent accessible (Figure 4b) and in environments where they have relatively unrestricted motion compared to other regions of Im9. The large RMSDs for these loops may therefore reflect mobility, and this may also account for the small number of NOEs observed for these regions (Figure 4d). Line widths of peptide NH resonances in $^{1}H-^{15}N$ HSQC spectra do not vary significantly along the peptide chain, showing that any differential mobility is not on a time scale to affect these. The origin of the large RMSDs in the 28−32 and 57−64 loops is currently under investigation with variant proteins produced by site-directed mutagenesis and relaxation time measurements.

The core of Im9 is stabilized by a large number of interactions involving hydrophobic side chains which, with the exception of some of the buried aromatic residues, are relatively well-defined. Figure 4c shows that the RMSDs of the side chains of Tyr10, Phe15, Phe40, and Phe83 are 1.0−1.8 Å. In addition to this, their respective H^{δ} and H^{ϵ} chemical shifts are degenerate. This indicates that these residues flip rapidly about their $C^{\beta}-C^{\gamma}$ bonds (Wüthrich, 1986), though the necessary structural fluctuations required to permit this are not fully reflected in the ensemble of structures shown in Figure 5. This situation has been observed for the majority of protein solution structures and indicates that an ensemble of NMR structures do not reflect all large scale amplitudes in proteins. On the basis of fractional surface accessibility calculations, with a value of <0.2 indicating removal from solvent (Figure 4b), the core residues include the four isoleucines, four of the five valines, three of the five leucines, the three phenylalanines, the tryptophan, two of the three histidines, and one of the three tyrosines. Most of these residues tend to be located on the interior faces of the helices.

Charged residues are located at the molecular surface, where they form favorable interactions with the solvent, but Lys4 and Asp9 appear to form a salt bridge. Although the long helices are generally amphipathic, there are some exceptions, most notably, the exposure of Ala25 on helix I and Val34 on helix II.

Fold of Im9. To the best of our knowledge, the Im9 topology has not been seen before. We base this assertion on a search we carried out of the protein folds deposited in the Brookhaven Protein Data Bank (558 protein chains at the time the search was made) with the FSSP program run using the Dali server at EMBL (Heidelberg) (Holm & Sander, 1993, 1994). The Im9 fold, which can be considered an antiparallel four-helix bundle in which one of the helices, helix III, has been terminated prematurely at the C-terminal end by a proline, residue 56, shows no similarity to that of the immunity protein of the RNase colicin E3 (Yajima *et al*., 1993). Osborne *et al*. (1994) noted that Im9 and the RNase inhibitor protein barstar (Guillet *et al*., 1993) exhibited some similarity in their distribution of secondary structural elements, but their tertiary structures are quite different. Nevertheless, it is interesting to note that barstar inhibits barnase by binding to it with a helix−loop segment (Guillet *et al*., 1993) and that helix II is involved in specificity determination of the Im9 interaction with the colicin DNases (Wallis *et al*., 1992a).

Structure−Function Relationships of Im9. Im9 forms a specific, stoichiometric complex with the endonuclease domain of Col E9 that completely inhibits the action of the enzyme through formation of one of the strongest noncovalent interprotein complexes reported in the literature (Wallis *et al*., 1995a). Previous genetic and biochemical studies (Wallis *et al*., 1992a) in combination with earlier structural work (Osborne *et al*., 1994) have suggested that helix II of Im9 is important for this interaction, particularly Val34 which is located on the surface of the protein. Helix II encompasses the so-called "specificity determining region", identified through homologous recombination experiments (Wallis *et al*., 1992a); position 34 was originally identified as an important determinant of specificity because substituting the valine of Im9 for an aspartic acid, the residue found at this position in Im8, yielded a chimeric protein that provided bacterial cells with significant biological protection toward the cytotoxic action of both the Col E9 and Col E8 toxins, whereas Im9 itself showed no cross-reactivity toward Col E8 in *in vivo* assays. Thus, position 34, and possibly other residues on the surface of helix II, are implicated in differentiating between the endonuclease domains of the different colicins.

Although the traditional view of immunity specificity is one in which each Im protein is totally specific for its target endonuclease, recent data on the E9 system have shown that this is not the case (Wallis *et al*., 1995b). E9 DNase can be bound by both the cognate (Im9) and noncognate (Im2, Im7, and Im8) immunity proteins, though the latter bind more weakly. However, the noncognate interprotein interactions still result in inhibition of the E9 DNase activity, suggesting that all four members of the Im family possess a common region on their surface that allows them to either directly bind to the active site of the enzyme or sterically block it. With this possibility in mind, it is of interest to consider how the distribution of helices and loops in Im9 (Figure 6) relates to the conservation of amino acid sequence in the immunity protein family. The N- and C-termini, which form extended strands, are highly conserved, but the short loop connecting helix I and helix II and the loop connecting helix III and helix IV contain variable amino acids. Of the four helices, all the amino acids of III and many of IV are conserved. Conversely helices I and II show little sequence conservation except for the N-terminal turn of helix I and the C-terminal turn of helix II. In looking for potential conserved regions lying close to helix II which could be involved in a common interaction with the E9 DNase domain, we found the structure of Im9 suggests two candidates: the N-terminus or helix III. Two lines of evidence suggest that it is helix III and not the N-terminus that is involved. Firstly, preliminary 2D NMR data using ^{15}N-labeled Im9 bound to unlabeled DNase show extensive chemical shift perturbations in the Im9 helix II and helix III regions but no changes in the N-terminal region (Osborne *et al*., unpublished data). Secondly, a peptide encompassing

the N-terminal region of Im9 does not affect the activity of the E9 DNase even at high concentrations of peptide relative to the DNase (Wallis *et al*., unpublished observations). Im9 is an acidic protein, its pI = 4.5, and kinetic analysis of its binding to the basic E9 DNase domain, whose pI > 10, has revealed that the rate of association is very salt dependent and in buffers of low ionic strength reaches a limiting value of 4×10^9 M^{-1} s^{-1} (Wallis *et al*., 1995a). These data indicate that there is an electrostatically steered collision between the two proteins to form an encounter complex which then undergoes a structural rearrangement to yield the final stable structure. This model implies that the DNase binding site of Im9, which is likely to include helices II and III, should be rich in acidic groups, and this is indeed the case. Glu30, Glu31, Glu32, Glu41, Asp45, and Asp51 are all part of this surface. By contrast, there is only one basic group in this region, Lys35. Mutagenesis studies, currently underway, are required to delineate further the interaction region on Im9.

ACKNOWLEDGMENT

We gratefully acknowledge Prof. K. Wüthrich for making available to us the program DIANA. We also thank Dr. R. Boetzel (UEA) for her assistance with the Ramachandran analysis.

SUPPORTING INFORMATION AVAILABLE

A virtually complete list of ^{1}H, ^{13}C, and ^{15}N assignments for Im9 (7 pages). Ordering information is available on any current masthead page.

REFERENCES

Chak, K.-F., & James, R. (1986) *J. Gen. Microbiol. 132*, 61−71.
Chak, K.-F., Kuo, W.-S., Lu, F.-M., & James, R. (1991) *J. Gen. Microbiol. 137*, 91−100.
Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 8172−8184.
Di Masi, D. R., White, D. C., Schnaitman, C. A., & Bradbeer, C. (1973) *J. Bacteriol. 115*, 506−513.
Eaton, T., & James, R. (1989) *Nucleic Acids Res. 17*, 1761−1761.
Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., & Feeney, J. (1990) *J. Magn. Reson. 90*, 420−425.
Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 99*, 201−207.
Grzesiek, S., & Bax, A. (1992b) *J. Am. Chem. Soc. 114*, 6291−6293.
Grzesiek, S., & Bax, A. (1993) *J. Biomol. NMR 3*, 185−204.
Guillet, V., Lapthorn, A., Hartley, R. W., & Maugen, Y. (1993) *Structure 1*, 165−176.
Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.
Holm, L., & Sander, C. (1993) *J. Mol. Biol. 223*, 123−138.
Holm, L., & Sander, C. (1994) *Nucleic Acids Res. 22*, 3600−3609.
James, R., Curtis, M. D., Wallis, R., Osborne, M. J., Kleanthous, C., & Moore, G. R. (1992) in *Bacteriocins, microcins & lantibiotics* (James, R., Lazdunski, C., & Pattus, F., Eds.) pp 181−201, NATO ASI Series H, Springer, Heidelberg.
Jardetzky, O., & Roberts, G. C. K. (1981) *NMR in Molecular Biology*, p 153, Academic Press, New York.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys 71*, 4546−4553.
Kay, L. E., Clore, G. M., Bax, A., & Gronenborn, A. M. (1990) *Science 249*, 411.
Koradi, R., Billeter, M., & Wüthrich, K. (1996) *J. Mol. Graphics* (in press).
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Lubienski, M. J., Bycroft, M., Freund, S. M. V., & Fresht, A. R. (1994) *Biochemistry 33*, 8866−8877.
Ludvigsen, S., & Poulsen, F. M. (1992) *J. Biolmol. NMR 2*, 227−233.
Luria, S. E., & Suit, J. L. (1987) in *Escherichia coli & Salmonella typhimurium, cellular & molecular biology* (Neidhardt, F. C., Ed.) Vol. 2, pp 1615−1624, American Society for Microbiology, Washington, DC.
Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259−281.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Marion, D., Driscoll, P,. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989a) *Biochemistry 28*, 6150−6156.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989b) *J. Magn. Reson. 85*, 393−399.
McCoy, M. A., & Mueller, L. (1992) *J. Am. Chem. Soc. 114*, 2108−2112.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
Osborne, M. J., Lian, L.-Y., Wallis, R., Reilly, A., James, R., Kleanthous, C., & Moore, G. R. (1994) *Biochemistry 33*, 12347−12355.
Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D., & Kay, L. E. (1994) *J. Magn. Reson. B103*, 197−201.
Redfield, A. G., & Kunz, S. D. (1975) *J. Magn. Reson. 19*, 250−254.
Schaller, K., & Nomura, M. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 3989−3993.
Serrano, L., Sancho, J., Hirshberg, M., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 544−559.
Toba, M., Masaki, H., & Ohta, T. (1988) *J. Bacteriol. 170*, 3237−3242.
Wallis, R., Moore, G. R., Kleanthous, C., & James, R. (1992a) *Eur. J. Biochem. 210*, 923−930.
Wallis, R., Reilly, A., Rowe, A., Moore, G. R., James, R., & Kleanthous, C. (1992b) *Eur. J. Biochem. 207*, 687−695.
Wallis, R., Reilly, A., Barnes, K., Abell, C., Campbell, D. G., Moore, G. R., James, R., & Kleanthous, C. (1994) *Eur. J. Biochem. 220*, 447-454.
Wallis, R., Moore, G. R., James, R., & Kleanthous, C. (1995a) *Biochemistry 34*, 13743−13750.
Wallis, R., Leung, K.-Y., Pommer, A. J., Videler, H., Moore, G. R., James, R., & Kleanthous, C. (1995b) *Biochemistry 34*, 13751−13759.
Wishart, D. S. & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
Wüthrich, K. (1986) in NMR of Proteins and Nucleic Acids, Wiley, New York.
Yajima, S., Muto, Y., Morikawa, S., Nakamura, H., Yokoyama, S., Masaki, H., & Uozumi, T. (1993) *FEBS Lett. 333*, 257−260.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405−410.
Zuiderweg, E. R. P., & Fesik, S. (1989) *Biochemistry 28*, 2387−2391.

BI960401K