Journal of Biomolecular NMR, **19:** 293–305, 2001.
KLUWER/ESCOM

NMR studies of the sporulation protein SpoIIAA: Implications for the regulation of the transcription factor σ^F in *Bacillus subtilis*

Helena Kovacs[a,*], David Comfort[b], Matthew Lord[c], Michael Yudkin[a], Iain D. Campbell[b] & Michael Nilges[a,**]

[a]*European Molecular Biology Laboratory, D-69012 Heidelberg, Germany;* [b]*Department of Biochemistry, South Parks Road, Oxford OX1 3QU, U.K.;* [c]*Microbiology Unit, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, U.K.*

Received 22 December 2000; Accepted 3 January 2001

Key words: automated assignment, NOE, structure calculation, tuberculosis

Abstract

SpoIIAA participates in a four-component mechanism for phosphorylation-dependent transcription control at the outset of sporulation. We report the refinement of the solution structure of SpoIIAA by using the automated iterative NOE assignment method ARIA. To complement the structural data, the protein dynamics were determined by measuring the T_1, T_2 and NOE of the backbone ^{15}N-nuclei. The refined structure permits a discussion of the structural features that are important for the function of SpoIIAA in the regulation of the sporulation sigma factor σ^F, and for homologous regulatory pathways present in *B. subtilis* and in other bacilli.

Introduction

The soil bacterium *Bacillus subtilis* responds to metabolic stress by entering a pathway of differential gene expression called sporulation, that is, a single cell divides into two cells in which different genes are expressed. The sporulation process culminates in the formation of an enduring cell type, endospore, and the lysis of the mother cell (Piggot and Coote, 1976; Losick et al., 1986; Errington, 1996). The temporally and spatially distinct activation of sporulation-specific genes is directed by transcription factors called sigma factors. The first transcription factor with compartmentalized activity is σ^F in the forespore (Losick and Stragier, 1992). The phosphorylatable protein called SpoIIAA – together with two other proteins, a specific protein kinase and a protein phosphatase – participates in the regulation of σ^F in an ATP/ADP dependent manner (Min et al., 1993; Alper et al., 1994; Diederich et al., 1994; Duncan et al., 1995). Thus the sporulating bacterium provides a prokaryotic model system for cell differentiation, where the cell develops into two different cell types, albeit with identical genomes. Although several transcriptional regulation pathways governing the sporulation process have been delineated in detail, structural information on the participant proteins is still scarce.

We present the structure refinement of the response regulator SpoIIAA by using the recently developed, fully automated iterative NOE-assignment procedure ARIA (Nilges and O'Donoghue, 1998). The overall fold of the dephosphorylated form of the 117-residue protein in solution has been reported earlier (Kovacs et al., 1998). SpoIIAA is a novel α/β fold devoid of long loops and unstructured regions. Several calculations were initially performed to test the applicability of the ARIA procedure for a protein of this size. Thus we could determine a high-resolution NMR structure by the ARIA method by using a comprehensive NOE data set (on average over 15 NOE constraints per residue) applied without any explicit assignment, employing the $^3J_{HN\alpha}$ coupling constants, and starting

*Present address: Bruker AG, Industriestr. 26, CH-8117 Fällanden, Switzerland.
**To whom correspondence should be addressed. Present address: Institut Pasteur, 25–28 rue du docteur Raux, F-75015 Paris, France. E-mail: nilges@pasteur.fr

from the previously determined overall fold. ARIA was also able to calculate the overall fold using the same data set but starting from a random template. The precision of the calculation could be enhanced by implementing ambiguous α-helical hydrogen bonds (i,i+4 or i,i+3) which are readily identified based on the primary NOE, $^{3}J_{HN\alpha}$ and $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift data. The use of the α-helical hydrogen bonds was, however, not indispensable for the success of the calculation.

The heteronuclear nuclear Overhauser effects (NOE), longitudinal (T_1) and transverse (T_2) relaxation rates of amide-^{15}N in the protein backbone were measured, thus complementing the structural data. The current NMR studies of SpoIIAA are a step towards elucidating – in terms of the four component proteins and the nucleotide – the structural basis for this distinct mechanism for regulation of differential gene transcription by σ^F and homologous pathways. Investigations of sequence homology (Park and Yudkin, 1997) and protein function (see, for instance, Duncan et al., 1996; Magnin et al., 1997) have previously indicated what the structural studies of SpoIIAA now confirm, that this regulatory mechanism has distinct features as compared to the known phospho-relayed signal transduction mechanisms.

SpoIIAA is classified into a novel fold family in the structural protein database (Hubbard et al., 1999), since the arrangement of the secondary structure elements in the SpoIIAA structure does not have a counterpart among the known protein folds. The primary sequences of SpoIIAA for *B. subtilis* and six other bacilli are highly (35 to 83% identity) homologous (Park and Yudkin, 1997). The sequences of two other *B. subtilis* proteins, RsbV and RsbS, are 32% and 22%, respectively, identical in sequence to SpoIIAA. These two proteins are known to participate in phosphorylation-dependent regulation of transcription in *B. subtilis* (Alper et al., 1996), with known functional and sequence – and thus presumably structural – homology to σ^F-regulation. Sequence homologues of the four sporulation-specific sigma factors have also been located in the complete genome of *Mycobacterium tuberculosis* (Cole et al., 1998). The possible sporulation sigma factor σ^F from *M. tuberculosis* has been isolated and in vitro studies suggest that its activation might be instrumental for the dormant state of this organism after infection (DeMaio et al., 1996). Our search in the *M. tuberculosis* genome with the SpoIIAA sequence confirmed the finding of Cole et al. (1998) of a protein sequence of comparative length and with 23% sequence identity to the SpoIIAA sequence from *B. subtilis* (M. Huynen and H. Kovacs, unpublished). Conserved in this alignment are, in particular, the phosphorylatable serine, residues close to it and several other key residues for the fold and function as discussed below; for instance, G20–D23 and G95–L96. The sequence homology suggests that this protein corresponds either to SpoIIAA in the σ^F regulation, or to RsbV or RsbS in the homologous σ^B regulatory pathway, although its sequence homology to the two latter ones is slightly lower, 18% and 15%. The genes for the proteins participating in the σ^F and σ^B regulation, respectively, are found in single operons in the *B. subtilis* genome (Kunst et al., 1997), whereas in the *M. tuberculosis* genome their homologues are shuffled (Cole et al., 1998), which complicates their identification.

Recently, statistically significant homology was reported between the SpoIIAA sequence with its bacterial homologues and the carboxy-terminal, cytoplasmic fragments from three human sequences belonging to the sulfate transporter family (Aravind and Koonin, 2000). The similarity involves 16 hydrophobic residues that are found in the core of the SpoIIAA structure, and G20 which is a conserved position in the alignment. Based on this homology, Aravind and Koonin define a fold superfamily, called the STAS-domain superfamily, which includes the bacterial anti-sigma factor antagonists (ASA) and the cytoplasmic domains of the sulfate transporter family (ST); the latter comprise a number of anion transporters both of eukaryotic and prokaryotic origin.

Materials and methods

Protein purification and ^{15}N-labeling

SpoIIAA from *Bacillus subtilis* was overproduced in *E. coli* by using the T7 RNA polymerase system. A detailed description of the purification procedure has been reported previously (Diederich et al., 1994; Kovacs et al., 1998). Methionine-1 was post-translationally cleaved yielding a 116-residue protein with an N-terminal serine. For the uniform ^{15}N-enrichment the cells were grown on minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source, the yield being 10–15 mg protein per litre. The maximal SpoIIAA concentration was about 1 mM in the NMR sample buffer which contained 25 mM NaCl, 25 mM K_2HPO_4, 1 mM dithiothreitol (DTT) and D_2O (5% v/v) at pH 6.4.

Nuclear magnetic resonance spectroscopy

The NMR experiments were performed on samples containing 1.0 mM unlabelled or 0.7 mM uniformly ^{15}N-labeled SpoIIAA protein at 25 °C. The spectra for the collection of distance constraints were recorded on a homebuilt 750 MHz spectrometer at Oxford University and the heteronuclear NOE, heteronuclear relaxation and HNHA-experiments were performed on a Bruker DRX 600 MHz spectrometer. All spectra were calibrated against a water frequency of 4.65 ppm at 25 °C. Spectral processing was done with FELIX software (FELIX 2.3, Molecular Simulations Inc.) and with NMRPipe (Delaglio et al., 1995) while the XEASY (Bartels et al., 1995) program was used for spectral analysis and determination of the signal intensities.

The resonance assignments of SpoIIAA have been reported earlier (Comfort, 1998; Kovacs et al., 1998). NOE cross peaks were collected from 2D-NOESY spectra in H_2O and D_2O of the unlabelled sample and a ^{15}N-correlated 3D-NOESY spectrum (ibid.). Additional 2D-NOESY spectra in H_2O were recorded with mixing times of 50, 75 or 125 ms at the 600 MHz ^{1}H field in order to monitor spin diffusion effects. The homonuclear two-dimensional NOESY spectra were typically recorded with a data size of 1024×512 complex points and $t_{2,max} = 102.4$ ms, $t_{1,max} = 51.2$ ms and with 48 scans per t_1-increment. The acquisition parameters for the three-dimensional ^{15}N-NOESY-HSQC experiment were: $t_{3,max} = 48.6$ ms (^{1}H, 512 complex points), $t_{2,max} = 14.2$ ms (^{1}H, 128 complex points) and $t_{1,max} = 8.4$ ms (^{15}N, 16 complex points). The ^{15}N-carrier was set to 116.0 ppm. The mixing time was 125.0 ms and 22 scans were collected per increment giving a total experiment time of about 70 h. For the coupling constants $^3J_{HN\alpha}$ a three-dimensional HNHA spectrum with water-flipback (Kuboniwa et al., 1994) was recorded on the ^{15}N-labeled sample. The sweep widths were 16.663 ppm in the acquisition, 10.414 ppm in the indirect proton and 28.155 ppm in the ^{15}N dimension. The size of the data matrix was 1024×128×64 complex points ($t_{3,max} = 51.2$ ms, $t_{2,max} = 20.5$ ms, $t_{1,max} = 37.37$ ms). The total mixing time was 40.4 ms. Eight scans were collected per increment and the total experimental time was 3.5 days.

For the determination of the ^{15}N longitudinal and transverse relaxation and ^{1}H-^{15}N steady-state NOE two-dimensional experiments utilizing coherence selection by pulsed field gradients were used (Farrow et al., 1994). The ^{15}N-correlated spectra consisted of 1024 complex points in the acquisition dimension with a spectral width of 16.66 ppm, and 128 complex points in the indirect dimension with a spectral width of 41.1 ppm, where the ^{15}N carrier was placed at 116.0 ppm. Thus $t_{2,max} = 51.3$ ms and $t_{1,max} = 51.2$ ms. For the T_1-measurements, eight spectra were recorded with relaxation delay T set to 10, 60, 100, 300, 500, 700, 1000 and 1200 ms. For T_2-measurements, eight spectra were recorded with the CPMG-type of relaxation delay T set to 16, 32, 48, 64, 80, 96, 112 and 128 ms. The equilibration delay between the scans was set to 1.2 s in the relaxation experiments and to 2.0 s in the heteronuclear NOE experiment. The T_1 and T_2 relaxation experiments were acquired with 32 scans per increment and the heteronuclear NOE experiment with 20 scans. The total measuring time was 23 h for the T_1-series, 26 h for the T_2-series and 15 h for each pair of a heteronuclear-NOE and a reference spectrum without NOE.

Collection of conformational constraints

The assignment of the ^{1}H, ^{15}N and ^{13}C resonance frequencies has been reported earlier (Kovacs et al., 1998). The cross peaks in the NOESY spectra were carefully peak-picked while comparing the presence and intensity of the signals at different mixing times to avoid peaks due to magnetization transfer through spin-diffusion. The NOE cross peaks were volume integrated by using routines in XEASY. The vicinal scalar coupling constants $^3J_{HN\alpha}$ were determined from the ratio of the diagonal and cross-peak intensities in the HNHA spectrum. The values were compared to those reported earlier, obtained by curve fitting to traces extracted from the HMQC-J spectrum. The previously reported 80 values were confirmed and 12 additional coupling constants were determined. The coupling constants were converted to ϕ dihedral angle constraints through the Karplus relation.

Automated iterative NOE assignment and structure calculation

The NMR-structure calculations were performed by using the fully automated iterative method for assigning NOE cross peaks, ARIA, Ambiguous Restraints for Iterative Assignment (Nilges and O'Donoghue, 1998), which is implemented in the software package XPLOR (Brünger, 1993). This strategy relies on the availability of a comprehensive assignment of the ^{1}H-resonance frequencies and frequency lists of integrated cross peaks with a negligible number of noise

peaks. In the present case three data sets were used: 2D-NOESY in H_2O and D_2O, and ^{15}N-correlated 3D-NOESY. At the outset the program performs a partial assignment, calibration, violation analysis and merging of the peak lists from different spectra. The resulting non-redundant set of NOE-based distance restraints is used in a simulated annealing calculation of 20 structures. Thereafter the NOE analysis is repeated, now based on the conformation of the seven lowest-energy structures from the calculation. This procedure is repeated iteratively eight times. The automated assignment of the NOE cross peaks is solely based on the chemical shifts of the protein protons, allowing for a frequency deviation of $+/-$ 0.015 ppm in the acquisition dimension and $+/-$ 0.030 ppm in the indirect proton dimension.

A unique characteristic of the method is the use of ambiguous distance restraints to treat degeneracy of several resonance frequencies. In the course of the iterative procedure this so-called dispersion degeneracy is resolved in the following way. The signal intensity of an ambiguous NOE corresponds to a distance **D** that is the sum of the contributions from all the possible assignments. Thus **D** is always shorter than any single contributing distance. In the analysis of the calculated structures an average distance is determined for each of the distances corresponding to a possible assignment. Then each contribution is estimated as a fraction of the sum of all contributions. A cutoff for accepting or rejecting possible assignments is defined by the user, such that the sum of the accepted contributions must fulfill the cutoff value. The cutoff (denoted by **p** in Table 1) is varied from one iteration to the next. In the beginning it is given a high value, 0.999, which means that few possible assignments, if any, are rejected. In the final iteration the cutoff is set to a low value, 0.80, which results in only distances of similar size being retained. If a single possible distance fulfills the cutoff, the NOE is assigned unambiguously to this distance. This approach of using the cutoff for the sum of the individual contributions instead of a straightforward distance cutoff as the criterion for accepting possible assignments, renders the NOE assignment a self-adjusting process, where the relative contributions rather than the actual distances are compared.

The calibration of cross-peak intensities against distances is done automatically in the beginning of every iteration. The assigned peaks corresponding to distances $<$3.5 Å in the calculated structures are included in the averaging in order to obtain a reference distance and a corresponding reference volume. Hence in the initial iterations the reference is mainly based on intra- and sequential connectivities, while successively an increasing number of medium- and long-range distances are included in the evaluation. However, our experience shows that the reference distances do not vary much and the calibration method is rather stable. Upper and lower bounds on a distance **d** are set to **d** $+/-$ $0.125*\mathbf{d}^2$.

Two ARIA calculations were performed starting either from a random template or the known fold. All the cross peaks were initially without assignment. Restraints that were systematically violated in the calculated structures with lowest energy were removed, also when they were assigned. These signals may be due to noise or spin diffusion. The output assignments from the latter run were analyzed for two purposes; to localize noise peaks and to identify the secondary structure elements. The excluded peaks were carefully checked and their presence and intensity in the NOESY spectra recorded with 50 and 75 ms mixing times was examined manually. In the final iteration 100 structures were calculated starting from the known fold. Forty structures with the lowest energy were further refined in a 9 Å shell of solvent water (Nilges et al., 1997) and 33 lowest-energy structures were analyzed. For the visual display and quantitative analysis of the calculated structures, the molecular graphics sofware MOLMOL (Koradi et al., 1996) was used. The quality of the calculated structures was inspected by using the PROCHECK software (Laskowski et al., 1996).

Analysis of the heteronuclear NOE and relaxation data

The signal intensities were determined by the maximum integration method in XEASY. Due to overlap, 8 signals were excluded from the analysis. Uncertainty due to noise in the measured peak heights can be approximated to $>$ 3% in the relaxation measurements and 5–8% in the two sets of heteronuclear NOE experiments, based on the baseline signal/noise level in the first spectra of the relaxation series and the S/N level in the spectra in the presence of NOE. The T_1 and T_2 values were obtained by non-linear least squares fitting of the peak intensities to an exponential function. The standard deviations in the curve fits were $<$ 10%. In the subsequent analysis the usual equations for the longitudinal and transverse relaxation time were used (Kay et al., 1989). These express T_1 and T_2 in terms of the spectral density functions and include contributions from the dipole–dipole and chemical shift

Table 1. Assignment statistics

Iteration	**p**	Violation tolerance (Å)	Number of distance constraints		
			ambiguous	unambiguous	total
1	0.999	2.0	1584	229	1744
2	0.99	1.0	1724	535	2259
3	0.98	0.5	1870	825	2695
4	0.96	0.25	1832	1319	3151
5	0.93	0.1	1390	1414	2804
6	0.90	0.25	1138	1600	2738
7	0.80	0.1	955	1655	2610
8	–	0.0	566	1751	2317
9[a]	–	566	1751	2317	

[a]Refinement of the final ensemble in explicit water.

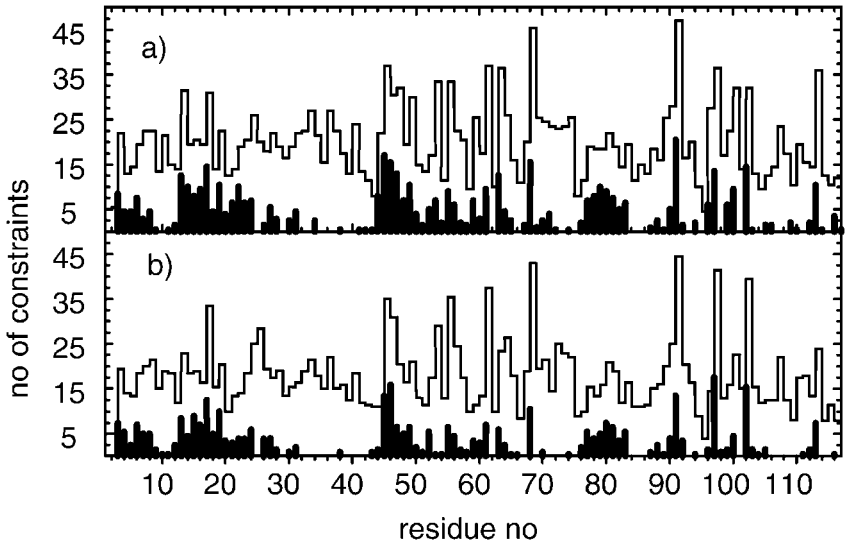

Figure 1. Number of NOEs per residue determined by (a) the automated routine and (b) manual assignment. Long-range connectivities are denoted by a solid bar and the total number by an open bar. From the ambiguous data in the automated calculation the first alternative assignment is included.

anisotropy relaxation mechanisms. For the spectral densities the simple form of $J(\omega_i) = S^2\tau_M/(1+\omega^2\tau_M^2)$ was found to be appropriate. This model assumes the overall tumbling to be isotropic small-step diffusion of a spherical rigid body, where the order parameter S^2 accounts for the initial fast-time-scale averaging of the reorientational correlation function. The reorientational correlation time τ_M was solved numerically for each residue from the T_1/T_2 ratio.

Results

The automated NOE assignment and structure calculation

The number of NOE-based distance constraints in the final ARIA iteration was 265 more than in the manual NOE-assignment reported earlier (Kovacs et al., 1998). It is noteworthy that 145 of the additional distance constraints are long-range (including ambiguous NOEs where the first alternative is a long-range connectivity). The distribution of the distance constraints per residue, obtained by the automated and manual assignments, is shown in Figures 1a and 1b. A conceptual difference between the two NMR-structure calculations is the use of ambiguous NOEs in the au-

Table 2. Summary of the constraints and the structure calculation

Conformational constraints		
NOE-based distance bounds:		
intra-residue $\|j-i\| = 0$	916	40%
sequential $\|j-i\| = 1$	598	26%
medium range $2 \leq \|j-i\| \leq 4$	309	13%
long-range $\|j-i\| > 5$	494	21%
total	2317	
unambiguous	1751	76%
ambiguous	566	24%
Dihedral angle constraints:		
$^3J_{HN\alpha}$ based backbone ϕ	90	
Quality of the calculated structures		
Violations:	33 best structures[a]	<NMR>[b]
distance bounds (> 0.3 Å)	1	0
dihedral angle constraints (> 5°)	0	0
Ramachandran analysis[c]		
residues in the most favoured regions	71.9%	
additional allowed regions	24.4%	
generously allowed regions	3.3%	
disallowed regions	0.4%	
Rmsd differences[d] in Å		
residues 2–117	backbone N, C^α and C'	all heavy atoms
	0.52 ± 0.09	0.95 ± 0.09
I helix (T26–L38)	0.22 ± 0.06	0.68 ± 0.11
II helix (S58–K72)	0.24 ± 0.07	0.58 ± 0.08
III helix (K88–S94)	0.11 ± 0.04	0.84 ± 0.17
IV helix (E106–T112)	0.25 ± 0.13	0.99 ± 0.33
I β-strand (G4–K10)	0.25 ± 0.06	0.71 ± 0.10
II β-strand (V13–T19)	0.17 ± 0.04	0.56 ± 0.11
III β-strand (H45–N49)	0.11 ± 0.04	0.34 ± 0.21
IV β-strand (E77–C81)	0.13 ± 0.05	0.58 ± 0.31

[a]Violated in 17 or more structures out of 33.
[b]The mean structure was obtained by averaging the coordinates of the 33 best structures after superposition of the N, C^α, C' backbone atoms in the structured regions comprising residues 2–113.
[c]Determined by the program PROCHECK (Laskowski et al., 1996).
[d]Root mean square deviation of the 33 best structures vs. the mean structure. Rmsd calculated for each region after the superposition of the indicated region.

tomated assignment routine. Twenty-eight percent of the NOE cross peak assignments in the final iteration remained ambiguous, cf. Table 1. In the manual assignment procedure the dispersion degeneracy was not taken into account, although it can be assumed to play a significant role for a protein of this size. In the manually assigned structure, the characteristic α-helical and cross β-strand hydrogen bonds were implemented explicitly, whereas no H-bonds were used in the current calculation. Although the number of $^3J_{HN\alpha}$ coupling constants used here was higher, the previous calculation incorporated additional constraints for backbone ϕ and ψ dihedrals based on the $^{13}C^\alpha$ chemical shift.

Structure of SpoIIAA

The solution structure of SpoIIAA resulting from the ARIA calculation is represented by the 33 lowest-energy conformers, energy-minimized in a shell of water; these have a backbone root mean square deviation (rmsd) of 0.52 +/− 0.09 Å to their mean. The rmsd is taken over the backbone atoms of all residues. The low value reflects the fact that the fold is compact and well-defined, lacking long unstructured termini

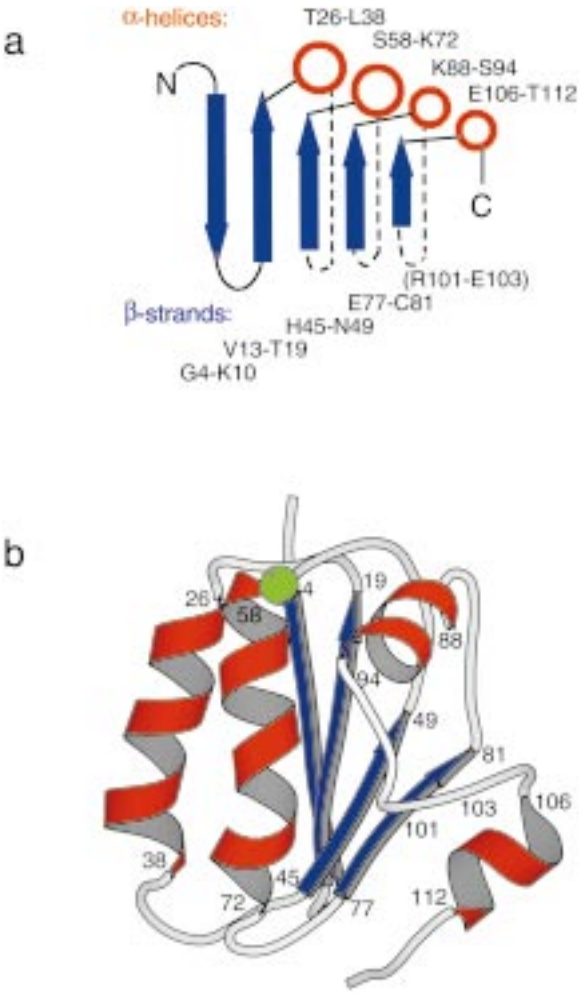

Figure 2. The topology of the α-helical and β-strand secondary structure elements in the SpoIIAA fold. A putative β-strand is indicated in parentheses, see text for details. (b) The three-dimensional fold of SpoIIAA, represented by the lowest energy structure from the ARIA calculation.

and loops. The data in Table 2 shows that the individual secondary structure elements are well-defined, and their arrangement is depicted schematically in Figure 2. The quality of the structure determination and the compactness of the protein architecture are illustrated by the uniform values of the per-residue rmsd in Figure 3. Compared to the previous structure of SpoIIAA (Kovacs et al., 1998), which was based on a manual NOE-assignment, the two structure determinations are of similar precision (backbone rmsd of 0.52 and 0.63 Å, respectively). However, the positions of the secondary structure elements differ slightly, in particular the position of the fourth α-helix. This reflects the increased number of long-range distance constraints

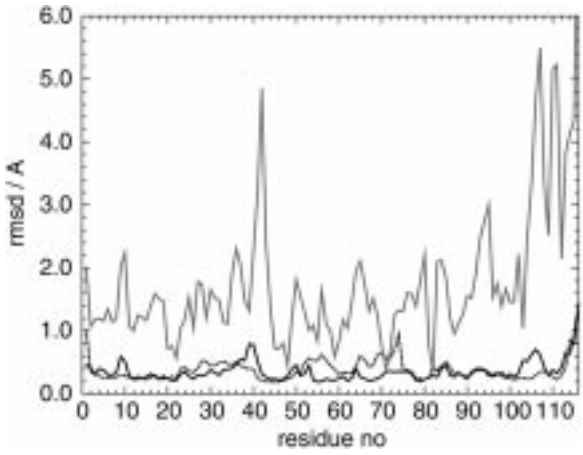

Figure 3. Per-residue rms deviations for the backbone atoms after a global superposition to the mean of the 10 best structures from the automated calculation (solid line) and the manually assigned calculation (dashed line), and after a superposition of these two average coordinates (gray line). Under the graph, arrows indicate β-strands, cylinders α-helical fragments and the position of the phosphorylatable serine-58 is marked by an asterisk.

in the C-terminus, cf. Figure 1. The superposition of the conformational spaces spanned by the 10 lowest-energy conformers from the automated and manual calculations is shown in Figure 4. Superposition of the average structures of the two structure bundles yields a backbone rmsd of 0.77 Å for the β-sheet in the protein core (including residues 4–10, 13–19, 45–49, 77–81) and a backbone rmsd value of 0.96 Å comprising the first, second and third α-helices (including residues 26–38, 58–72, 88–94), whereas the simultaneous superposition of all α-helical fragments and β-strands yields an rmsd of 1.80 Å. The differing regions can be identified through the superposition of the mean coordinates of the two structure bundles; the per-residue rmsd values after such a superposition are presented in Figure 4. The ARIA-calculated structure has fewer residual violations and no backbone torsion angles in disallowed regions.

SpoIIAA has an α/β-type fold that lacks known close structural homologues in the structure database of the Protein Data Bank (Kovacs et al., 1998). Figure 2 shows the alternating order of the β-strands and α-helices in the SpoIIAA-fold, the possible fifth β-strand is indicated in parentheses. The central pleated β-sheet consists of four β-strands with the first one in an anti-parallel orientation to the rest. The experimental data neither confirm nor contradict the presence of an additional, short parallel β-strand comprising residues R101–E103 next to the fourth strand

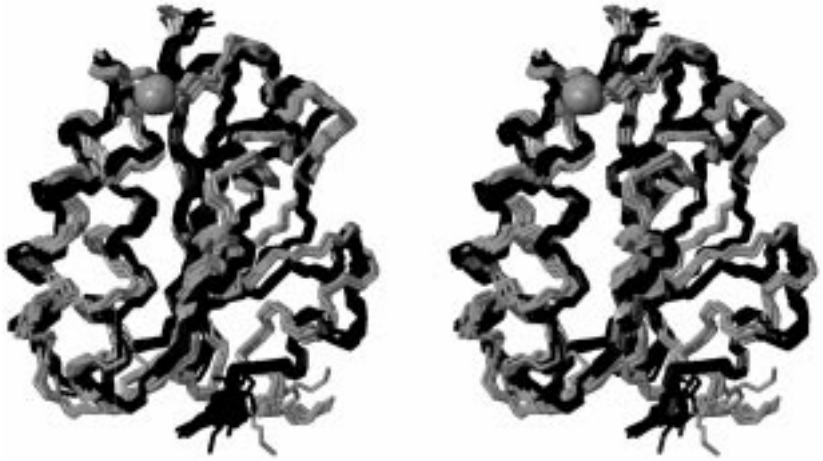

Figure 4. SpoIIAA fold. Superposition of the automatically determined structure of SpoIIAA (dark colour) and the structure obtained by manual interpretation of the spectra (light colour). Ten lowest energy conformers of each structure were included in the global superposition (see Table 2, footnote). Image generated by MOLMOL (Koradi et al., 1996). The position of the phosphorylatable serine-58 is indicated by a sphere.

of residues E77–C81. Residues R101–E103 show the characteristic sequential $H^{\alpha}H^{N}$ distance of 2.2 Å. However, only two cross-strand distance constraints are found between the HN and H^{α} atoms, involving residues V80 and F102, and C81 and E103, respectively. Instead, there are numerous side-chain contacts between the involved residues. The α-helices are four in number. The first two helices are positioned on one side of the β-sheet and have numerous NOE-contacts to one another. The two shorter helices are located near the C-terminal edge of the central β-structure. The position of the last α-helix comprising residues 106–112 is defined by several distance constraints in the current structure.

Between the second and third helices there is a region devoid of charged or polar residues, but exposing a number of hydrophobic residues. This is illustrated in Figure 5. The electrostatic surface next to the phosphorylatable serine-58 is shown in Figure 5A with the positions of the charged residues indicated. In Figure 5B the exposed hydrophobic residues are mapped onto the van der Waals surface, color-coded and labeled. Figure 5C displays the solvent-accessible surface of the protein and makes visible the location of this exposed hydrophobic region between the second and the third α-helices. Certain residues adjacent to this region have been shown by mutagenesis to be important for the protein function. For instance, although the mutant proteins G62D and G95D retain their capacity to make non-covalent complexes with the kinase SpoIIAB and ADP (Barilla et al., 1999), they differ from the wild type. The mutation G95A does not affect the phosphorylation of SpoIIAA but abolishes function by impairing the hydrolysis by the phosphatase SpoIIE. The mutant G62D is phosphorylated, albeit slowly, and is completely unresponsive to SpoIIE (ibid.). Affinity chromatography indicates that the SpoIIAA–SpoIIAB interaction is of hydrophobic character (Duncan et al., 1996). The exposure of the hydrophobic residues illustrated in Figure 5 is likely to be significant for the SpoIIAA interaction with the kinase and the phosphatase.

Dynamics data

The longitudinal and transverse relaxation rates and heteronuclear NOEs are sensitive to motions on different time scales. The transverse relaxation rate T_2 is also affected by conformational exchange processes, provided they take place at a faster rate than the length of the refocusing delay in the CPMG scheme, which, in the present case, was set to 450 μs. The T_1 and T_2 relaxation times and the heteronuclear NOE of the backbone amide ^{15}N-nuclei depend on the dipolar effect of the directly bound amide-proton and the chemical shift anisotropy of the ^{1}H-^{15}N bond. At the ^{15}N-resonance frequency of 60.8 MHz the contribution to the ^{15}N-T_1 can be estimated to be 76% from the dipolar effect and 24% from the chemical shift anisotropy effect. The relaxation times and heteronuclear NOE for SpoIIAA at 14.1 T are presented in Figure 6.

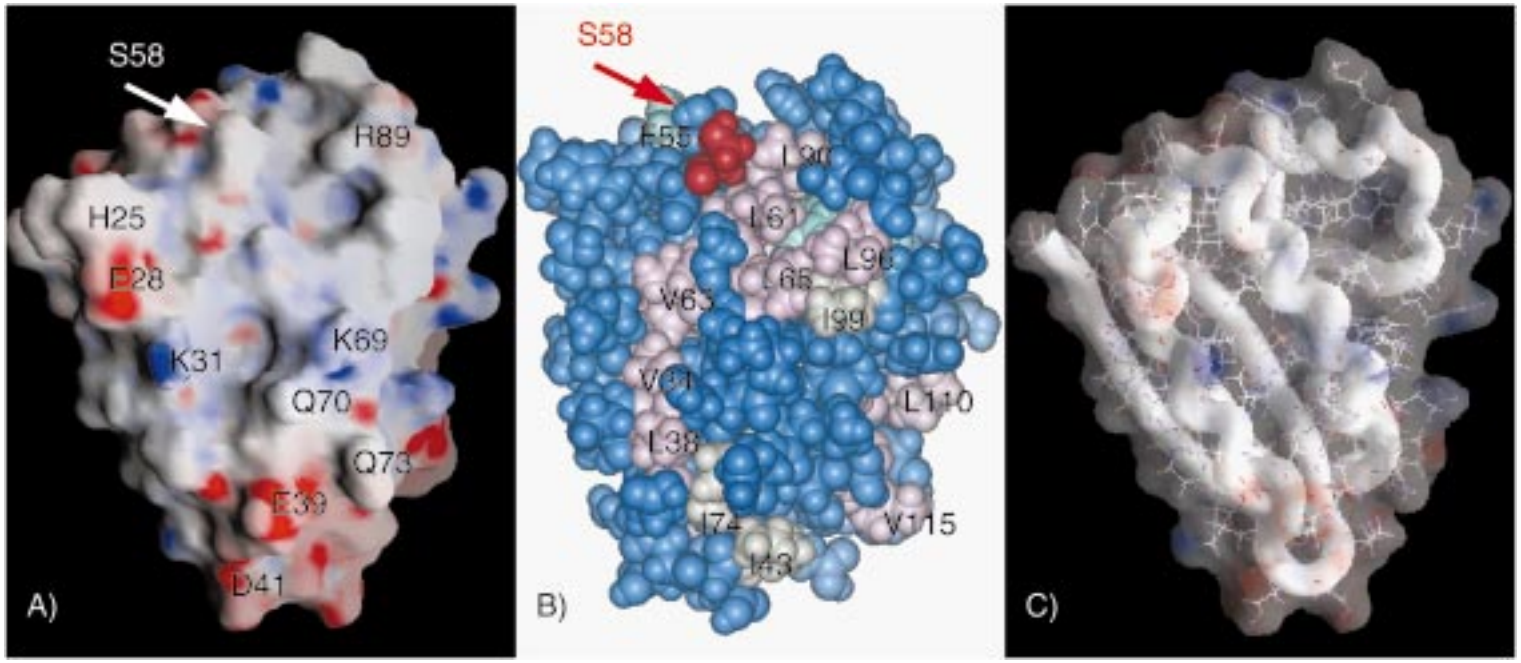

Figure 5. Surface properties of SpoIIAA. (A) The electrostatic surface of SpoIIAA, calculated with the program GRASP (Nicholls and Honig, 1991) is shown. (B) A hydrophobic solvent-exposed region is illustrated by color-coding on the van der Waals atomic spheres: valine – pink, leucine, isoleucine – beige, S58 – red. The figure was drawn by InsightII (Molecular Simulations Inc.). (C) The secondary structure elements and the solvent accessible surface are displayed.

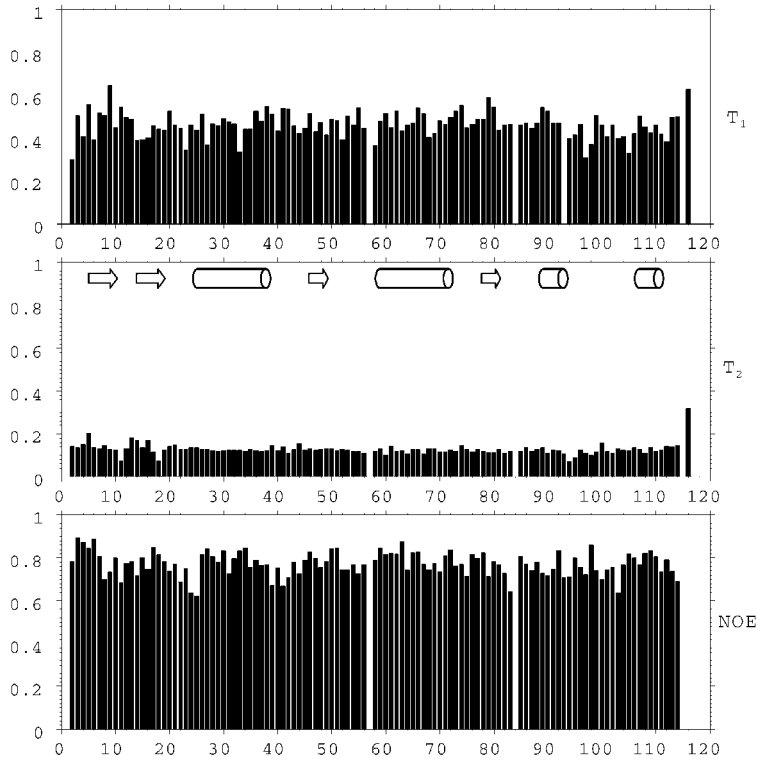

Figure 6. The heteronuclear NOE and the longitudinal and transverse relaxation times T_1 and T_2 of the backbone ^{15}N-nuclei plotted as a function of the residue number. Positions of β-strands and α-helices are indicated by arrows and cylinders, respectively.

The average T_1 and T_2 values are 0.46 +/− 0.06 s and 0.13 +/− 0.02 s, respectively. The values for the individual residues are uniform within the limits of the experimental error; neither T_1 nor T_2 values show a systematic deviation involving a number of neighboring residues. The overall correlation time, τ_M, assuming isotropic reorientation of a spherical rigid body can be estimated from the T_1/T_2 ratios of the individual ^{15}N-nuclei (Kay et al., 1989). The assumption that SpoIIAA tumbles as a rigid body is also justified by the NOE-values, which indicate that the internal motions are negligible, their contribution to T_1 being about 2.5%. The approximation of isotropic reorientation is also applied here, although the three components of inertia of the SpoIIAA structure are 1.0:1.2:1.4. The average value of τ_M, 5.02 +/− 0.06 ns, was obtained from the T_1/T_2 ratios, excluding eight overlapping and three mobile terminal residues. This value is in good agreement with the values determined for proteins of similar size (Farrow et al., 1994).

An average ^{15}N-NOE value of 0.77 was obtained with a standard deviation of 0.05 from two independent measurements. The averages are taken over 105 resolved signals and exclude the terminal residues 116A, 117S and S2. For the two latter residues negative signal intensities were observed in the presence of ^{1}H-saturation, indicating considerable flexibility in the termini. Assuming that the protein rotates as a spherical rigid body in the regime $\omega_X\tau_M \gg 1$, the maximum ^{15}N-NOE would be 0.82, taking into consideration both the ^{1}H-^{15}N dipolar interaction and the chemical shift anisotropy (CSA) relaxation mechanisms. Thus, in SpoIIAA the internal motions contribute little to the measured ^{15}N-NOE. The backbone dynamics of SpoIIAA thus support the presence of a rigid, compact fold with limited conformational fluctuations.

Discussion

The high-resolution NMR-structure for the 117-residue protein SpoIIAA was obtained by the automated NOE assignment procedure ARIA. Compared to a manual assignment the ARIA routine localized a number of additional NOE connectivities, including several long-distance constraints, which are important

for the definition of the tertiary fold of the protein. One fourth of the NOE cross peaks consisted of several contributions due to dispersion degeneracy, a feature that is difficult to account for in a manual spectral assignment, although it is of considerable consequence for structure determination of proteins of current size and larger. It became apparent that the size of the protein is close to the limit for the present ARIA protocols for completely automated NOE assignment and high-resolution structure calculation, when the NOE data arise predominantly from homonuclear NOE spectra. ARIA has been used earlier in the structure determination of globular proteins of similar size such as the PH domain of 106 residues (Macias et al., 1994).

$^3J_{HN\alpha}$ coupling constants were necessary for the formation of the correct fold in ARIA calculations (data not shown) that started from extended random conformations. It appears that the dihedral angle constraints are indispensable for the correct formation of the α-helices. The correct fold could also be achieved in the ARIA runs starting from a random fold when using α-helical hydrogen bonds (allowing for (i,i+3) or (i,i+4) ambiguity), but not when using solely the NOE data. In SpoIIAA the α-helical hydrogen bonds are substantiated by ample data indicative of helical secondary structure; that is, characteristic NOE-patterns, small $^3J_{HN\alpha}$ coupling constants and the deviation of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts from the random coil values. The calculation starting from the random conformation converged very well to the correct fold when both $^3J_{HN\alpha}$ coupling constants and α-helical hydrogen bonds were used.

The structure of SpoIIAA corroborates the previous evidence that the regulatory mechanism of σ^F has unique features and significantly differs, for instance, from the so-called two-component phospho-relay mechanism. Also this mechanism was initially detected in *B. subtilis*, where it activates the sporulation transcription regulator Spo0A in the preseptational cell. Sequence homology indicates that the two-component phospho-relay is a frequently found signal transduction mechanism in procaryotes (Appleby et al., 1996) and recently its presence in eucaryotes has been pointed out (Posas et al., 1996). Although the σ^F-mechanism, in analogy to the two-component phospho-relay, also involves a sensor kinase and an effector (the transcription factor that is regulated), it does not contain a multi-step phospho-transfer pathway. Instead, at the onset of sporulation, the response regulator SpoIIAA circulates between the protein kinase (and anti-sigma-factor) and the protein phosphatase while σ^F is free to exert transcriptional regulation (Magnin et al., 1997).

The structure of Spo0F, the response regulator in the Spo0A control system, has been determined by NMR (Feher et al., 1997), as well as that of a close homologue, the chemotaxis protein CheY (Volz and Matsumura, 1991). Recently, the complex of the CheY binding domain of the sensor kinase CheA and CheY was published (Welch et al., 1998), where the binding domain functions as an activator domain. The complex structure indicates that in preparation for phosphorylation the structure of CheY undergoes a cascade of conformational changes that are propagated through the protein. NMR studies of the Spo0F structure and dynamics have given similar information (Feher et al., 1998). The Spo0F on one hand, and SpoIIAA on the other, appear to be functionally akin, since both are phosphorylatable stress response regulators in bacteria, and have α/β folds of similar size. The folds, however, differ. In particular, the location of the phosphorylation site is dissimilar, the phospho-serine in SpoIIAA being located in the N-terminal end of an α-helix and the phospho-asparagine in Spo0F and CheY being on a β-strand. Furthermore, the kinase SpoIIAB lacks an activation domain analogous to the one in the Spo0A regulation system. Taken together, the present structural and dynamic data from NMR show that SpoIIAA has a uniformly well-defined, compact and rigid structure, unlike the Spo0F/CheY structure. The number of NOEs and backbone rmsds per residue are practically uniform throughout the sequence and the observed backbone-^{15}N NOE, T_1 and T_2-values do not indicate local flexibility. Further, the phosphorylation site in SpoIIAA is on the surface of the protein, in contrast to that of Spo0F or CheY.

The phosphorylation site in SpoIIAA is S58 positioned at the N-terminus of the second α-helix. This is a favourable location for a phospho-group as it is stabilized by the helix dipole and by interactions with the exposed backbone NH-groups. S58 is also within 10 Å of the N-terminal end of the first α-helix. In general, the observed effects upon phosphorylation can be divided into two categories (Johnson and O'Reilly, 1996), an allosteric conformational change propagated through the molecule, as found in Spo0F and CheY, or a direct, local steric and/or electrostatic hindrance. An example of the latter is the serine-phosphorylation of the histidine-containing protein HPr, which is a component of the phosphoenolpuryvate:sugar phosphotransferase system in *B. subtilis*. The NMR studies of HPr show that the phosphorylated form is identical

to the non-phosphorylated form, apart from very localized effects around the phospho-site (Pullen et al., 1995). Similar to SpoIIAA, HPr is phosphorylated at a serine situated on the N-terminus of an α-helix. Comparisons were made to three other known structures with phospho-serines at a similar location and the conclusion was drawn that phosphorylation of a serine at the N-terminal end of an α-helix is not likely to result in a conformational change (Pullen et al., 1995).

Concluding remarks

The refined structure of SpoIIAA contributes to the structural characterization of the nucleotide binding mode of SpoIIAA, which appears to be different from the known nucleotide interaction sites. The structure and dynamics of SpoIIAA confirm that the SpoIIAA fold is dissimilar to its functional homologues CheY and Spo0F, the response regulators of the bacterial phospho-relay mechanism which have been well studied on the molecular level. Both the structural and dynamic data about SpoIIAA indicate restricted internal mobility. This, combined with what is known of other proteins with a phospho-serine at a similar location, suggests a local electrostatic effect at the phosphorylation site rather than a major conformational change. The SpoIIAA structure has further significance since, based on sequence homology, it represents a protein fold superfamily comprising a number of bacterial phospho-regulators and possibly also a family of anion transporters of procaryotic and eucaryotic origin.

References

Alper, S., Duncan, L. and Losick, R. (1994) *Cell*, **77**, 195–205.

Appleby, J.L., Parkinson, J.S. and Bourret, R.B. (1996) *Cell*, **86**, 845–848.

Aravind, L. and Koonin, E.V. (2000) *Curr. Biol.*, **10**, R53–R55.

Barilla, D., Lucet, I., Kuhlmann, A. and Yudkin, M.D. (1999) *J. Bacteriol.*, **181**, 3860–3863.

Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **6**, 1–10.

Brünger, A.T. (1993) *X-PLOR version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Challoner-Courtney, I.J. and Yudkin, M. (1993) *J. Bacteriol.*, **175**, 5636–5641.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K. et al. (1998) *Nature*, **393**, 537–544.

Comfort, D. (1998) Doctoral Thesis, University of Oxford.

Delaglio, F., Grzesiek, S., Vuister, G.W., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

DeMaio, J., Zhang, Y., Ko, C., Young, D.B. and Bishai, W.R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 2790–2794.

Diederich, B., Wilkinson, J.F., Magnin, T., Najafi, S.M.A., Errington, J. and Yudkin, M.D. (1994) *Genes Dev.*, **8**, 2653–2663.

Duncan, L., Alper, S., Arigoni, F., Losick, R. and Stragier, P. (1995) *Science*, **270**, 641–644.

Duncan, L., Alper, S. and Losick, R. (1996) *J. Mol. Biol.*, **260**, 147–164.

Errington, J. (1996) *Trends Gen.*, **12**, 31–34.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) *Biochemistry*, **33**, 5984–6003.

Feher, V.A., Zapf, J.W., Hoch, J.A., Whiteley, J.M., McIntosh, L.P., Rance, M., Skelton, N.J., Dahlquist, F.W. and Cavanagh, J. (1997) *Biochemistry*, **36**, 10015–10025.

Feher, V.A., Tzeng, Y.-L., Hoch, J.A. and Cavanagh, J. (1998) *FEBS Lett.*, **425**, 1–6.

Hubbard, T.J.P., Ailey, B., Brenner, S.E., Murzin, A.G. and Chothia, C. (1999) *Nuleic Acids Res.*, **27**, 254–256.

Johnson, L.N. and O'Reilly, M. (1996) *Curr. Opin. Struct. Biol.*, **6**, 762–769.

Kay, L.E., Torchia, D.A. and Bax, A. (1989) *Biochemistry*, **28**, 8972–8979.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graphics*, **14**, 51–55.

Kovacs, H., Comfort, D., Lord, M., Campbell, I.D. and Yudkin, M.D. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 5067–5071.

Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) *J. Biomol. NMR*, **4**, 871–878.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G. et al. (1997) *Nature* **390**, 249–256.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Moss, D.S., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Lord, M. (1998) Doctoral Thesis, University of Oxford.

Losick, R., Youngman, P. and Piggot, P.J. (1986) *Annu. Rev. Genet.*, **20**, 625–669.

Losick, R. and Stragier, P. (1992) *Nature*, **355**, 601–604.

Macias, M.J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M. and Oschkinat, H. (1994) *Nature*, **369,** 675–677.

Magnin, T., Lord, M. and Yudkin, M.D. (1997) *J. Bacteriol.*, **179**, 3922–3927.

Min, K.-T., Hilditch, C.M., Diederich, B., Errington, J. and Yudkin, M.D. (1993) *Cell*, **74**, 735–742.

Nicholls, A. and Honig, B. (1991) *J. Comput. Chem.*, **12**, 435–445.

Nilges, M., Macias, M., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.

Nilges, M. and O'Donoghue, S.I. (1998) *Progr. NMR Spectrosc.*, **32**, 107–139.

Park, S.G. and Yudkin, M.D. (1997) *Gene*, **194**, 25–33.

Piggot, P.J. and Coote, J.G. (1976) *Bacteriol. Rev.*, **40**, 908–962.

Posas, F., Wurgler-Murphy, S.M., Maeda, T., Witten, E.A., Thai, T.C. and Saito, H. (1997) *Cell*, **86**, 865–875.

Pullen, K., Rajagopal, P., Branchini, B.R., Huffine, M.E., Reizer, J., Saier Jr. M.H., Scholtz, J.M. and Klevit, R.E. (1995) *Protein Sci.*, **4**, 2478–2486.

Volz, K. and Matsumara, P. (1991) *J. Biol. Chem.*, **266**, 15511–15519.

Welch, M., Chinardet, N., Mourey, L., Birck, C. and Samama, J.-P. (1998) *Nat. Struct. Biol.*, **5**, 25–29.