Solution Structure of the Transcriptional Activation Domain of the Bacteriophage T4 Protein, MotA[†,‡]

Ning Li,[§] Weixing Zhang,[§] Stephen W. White,*[,§,||] and Richard W. Kriwacki*[,§,||]

Department of Structural Biology, St. Jude Children's Research Hospital, 332 North Lauderdale Street, Memphis, Tennessee 38105, and Department of Molecular Sciences, University of Tennessee Health Sciences Center, Memphis, Tennessee 38163

Received December 13, 2000; Revised Manuscript Received February 7, 2001

ABSTRACT: Bacteriophage T4 encodes a transcription factor, MotA, that binds to the −30 region of middle-mode promoters and activates transcription by host RNA polymerase. The crystal structure of the N-terminal domain of MotA (MotNF) revealed a six-helix domain in which the two C-terminal α-helices mediate the formation of a dimer via a coiled-coil motif and hydrophobic interactions. This structure suggested that full-length MotA binds DNA as a dimer, but subsequent biochemical results have shown that a monomeric form of MotA binds DNA. In this study, gel filtration chromatography, dynamic light scattering, and NMR-based diffusion measurements show conclusively that MotNF is a monomer, and not a dimer, in solution. In addition, we have determined the monomeric solution structure of MotNF using NMR spectroscopy, and have compared this with the dimer structure observed in crystals. The core of the protein assumes the same helical conformation in solution and in crystals, but important differences are observed at the extreme C-terminus. In solution, helix α5 is followed by five disordered residues that probably link the N-terminal and C-terminal domains of MotA. In crystals, helix α5 forms the dimer interface and is followed by a short sixth helix that further stabilizes the dimer configuration. The solution structure of MotNF supports the conclusion that MotA functions as a monomer, and suggests that the existence of the sixth helix in crystals is a consequence of crystal packing. Our work highlights the importance of investigating protein structures in both crystals and solution to fully understand biomolecular structure and to accurately deduce relationships between structure and function.

Crystal structures sometimes reveal unexpected multimeric states, and it can be difficult to assess whether these states have functional relevance in vivo. It is important to characterize these states because they can have implications for how the biological system operates at the molecular level. An example is the class II major histocompatibility (MHCII) molecule, where it is still unclear if the crystallographic dimer of dimers is physiologically relevant (*1*). Recently, we encountered this problem in the analysis of the MotA transcription factor from bacteriophage T4 (*2*). The structure of the N-terminal domain revealed an apparently stable and physiologically relevant dimer, but subsequent biochemical experiments do not support this multimeric state.

The MotA[1] transcription factor is specifically required for the activation of T4 middle genes during its infection of *Escherichia coli*. Early genes are expressed from promoters that mimic those of the host, and contain the typical −10 and −35 regions (*3*). Late genes have T4-specific promoters that are recognized by a modified host RNA polymerase in which the σ^{70} factor has been replaced with T4-encoded factor gp55 (*4*). The middle gene promoters are a compromise between these extremes (*5*). They resemble the host promoters in having a standard −10 sequence, but have a unique −30 sequence and are transcribed by a T4-modified host RNA polymerase. Preferential recognition of the T4-specific −30 sequence over the host −35 sequence is effected by two T4-encoded proteins, MotA and AsiA, that form part of the middle-mode transcription complex. Genetic and biochemical experiments have shown that only the T4-modified host polymerase, MotA, and AsiA are required for transcription from T4 middle promoters (*6*−*11*).

The precise interactions that occur between the polymerase, MotA, and AsiA at the middle promoter are still being investigated (*12*), but their individual roles have now been established. AsiA binds specifically to the σ^{70} factor (*13*, *14*), and it appears to promote a conformational change

[†] This work was supported by National Institutes of Health Cancer Center (CORE) Support Grant CA 21765 and the American Lebanese Syrian Associated Charities.

[‡] The coordinates for the 20 lowest-energy conformers of MotNF have been deposited in the Brookhaven Protein Data Bank (entry 1I1S).

* To whom correspondence should be addressed. R.W.K.: e-mail, Richard.Kriwacki@stjude.org; phone, (901) 495-3290; fax, (901) 495-3032. S.W.W.: e-mail, Stephen.White@stjude.org; phone, (901) 495-3040; fax, (901) 495-3032.

[§] St. Jude Children's Research Hospital.

[||] University of Tennessee Health Sciences Center.

[1] Abbreviations: AsiA, anti-sigma factor; DTT, dithiothreitol; HSQC, heteronuclear single-quantum coherence; IPTG, isopropyl β-D-thiogalactoside; MotA, modifier of transcription; MotCF, MotA C-terminal fragment; MotNF, MotA N-terminal fragment; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PEI, polyethylenimine; PMSF, phenylmethanesulfonyl fluoride; TAD, torsional angle dynamics; TALOS, torsion angle likelihood obtained from shift and sequence similarity; Tris, tris(hydroxymethyl)aminomethane; rmsd, root-mean-square deviation.

10.1021/bi0028284 CCC: $20.00
Published on Web 03/17/2001

that prevents recognition of the host −35 sequence (*15*). MotA is a specific DNA-binding protein that recognizes the −30 "mot box" sequence and that binds the AsiA−σ^{70} complex (*8*, *9*, *16*, *17*). The AsiA−σ^{70} complex and gp55, the virus-encoded σ^{70}-like factor, compete for the core polymerase complex, leading alternatively to transcription of middle-mode and late-mode viral genes (*18*). Together, the AsiA−σ^{70} complex and MotA recruit the host RNA polymerase to initiate from middle-mode promoters (*12*). We have shown that MotA comprises an N-terminal domain and a C-terminal domain, MotNF and MotCF, respectively, that can readily be separated proteolytically (*19*). Each has been independently expressed and purified, and we have determined the crystal structure of MotNF (*2*) and a preliminary NMR solution structure of MotCF (*20*). Mutagenesis experiments with MotNF (*2*, *21*) and gel mobility shift assays with MotCF (unpublished data) confirm that the former is involved in protein−protein interactions, and that the latter binds DNA.

The MotNF crystal structure reveals a small six-helix domain in which the two C-terminal α-helices mediate the formation of a dimer via a coiled-coil motif and hydrophobic interactions (*2*). This structure suggested that MotA binds DNA as a dimer, like many other transcription factors (*22*, *23*). However, recent studies of the DNA binding properties of MotA have shown that the cognate complex only involves a single molecule of MotA (*24*). In light of this work, and its important implications for the operation of middle-mode transcription in T4, we embarked on a detailed analysis of the solution properties of MotNF, including a complete determination of its solution structure using NMR spectroscopy to accompany the crystallographic analysis. The important conclusion of this work is that the dimeric structure of MotNF within the crystal is indeed an artifact that appears to be the result of truncating the full-length protein.

EXPERIMENTAL PROCEDURES

Expression and Purification. The N-terminal domain of MotA (MotNF, amino acids 1−96) was subcloned into bacterial expression vector pET3C (Novagen), and MotNF was expressed in BL21(DE3) cells. Uniformly ^{15}N- or ^{15}N- and ^{13}C-labeled MotNF was obtained by growing BL21(DE3) cells transformed with the MotNF plasmid in a defined medium (*25*) containing [^{13}C]glucose and/or [^{15}N]ammonium chloride, as appropriate. Routinely, cells (4 L) were grown at 37 °C until an absorbance at 600 nm of 0.6 was reached. Cultures were induced with 1 mM IPTG and grown to saturation. Cells were collected by centrifugation and lysed by treatment with lysozyme in lysis buffer [350 mM potassium phosphate (pH 7.6), 1 mM PMSF, 1% PEI, and 1 mM DTT]. The cell suspension was centrifuged, and ammonium sulfate was added to the supernatant to 80% saturation. The precipitated proteins were collected by centrifugation and dissolved in 25 mL of 50 mM KPO_4 (pH 7.6) and dialyzed against the same buffer overnight at 4 °C. The dialysate was passed through a 50 mL DEAE-Sepharose column (Amersham Pharmacia, Piscataway, NJ); MotNF is found in the flow-through fraction. This fraction was dialyzed against 20 mM Tris (pH 8.5) overnight at 4 °C. The dialysate was then loaded onto a 50 mL Q-Sepharose (Amersham Pharmacia) column equilibrated with the same buffer, and MotNF was eluted at 0.16 M NaCl using a salt gradient of 0.0 to 1.0 M over 5 column volumes. Fractions containing MotNF were pooled and concentrated by ultrafiltration using an Amicon cell with a 3 kDa cutoff membrane (Millipore, Inc., Bedford, MA). In the final step, MotNF was purified using a HiLoad 16/60 Superdex 75 gel filtration column (Amersham Pharmacia) equilibrated with 200 mM potassium phosphate buffer (pH 6.5).

Gel Filtration Analysis. Gel filtration chromatography was used to determine the oligomeric state of MotNF in solution. The elution times of molecular mass standards (Bio-Rad Laboratories, Inc., Hercules, CA) were first determined for a Superdex 75 column. The column was run in 200 mM potassium phosphate (pH 6.5) with a flow rate of 1.5 mL/min, and the sample was loaded in 200 μL. The ratios of the elution volumes to the void volume (V_e/V_0) for the standards were plotted against the natural log of molecular masses to establish a calibration curve. The elution volume of the 600 000 Da standard was taken to be the void volume. The elution time of MotNF was then determined under identical conditions, and the apparent molecular mass of MotNF in solution was calculated from the calibration curve.

Light Scattering Methods. Dynamic light scattering was used to estimate the hydrodynamic radius and MW of MotNF on a DynaPro-800WIN (Protein Solutions, Charlottesville, VA) molecular size detector, with appropriate viscosity and refractive index corrections. Light scattering results were obtained using a 1.5 mM MotNF sample prepared in 200 mM potassium phosphate (pH 6.5).

NMR Sample Preparation. The MotNF samples from the final purification step were first exchanged into 200 mM potassium phosphate buffer with 10% D_2O and 0.02% sodium azide (pH 6.5) and then concentrated to ∼2.0 mM in an Amicon ultrafiltration cell. Approximately 300 μL samples were transferred to Shigemi micro-cell NMR tubes (Shigemi, Inc., Allison Park, PA) for NMR spectroscopic studies.

NMR Spectroscopy. All NMR spectra were acquired at 28 °C (except diffusion experiments which were carried out at 25 °C) on a Varian Inova 600 MHz spectrometer (Varian, Inc., Palo Alto, CA) using a 5 mm triple-resonance probe equipped with xyz pulsed magnetic field gradients (PFGs). In general, well-established pulse sequences provided by Varian in the Protein Pack suite were used to acquire two- and three-dimensional NMR data. Experiments not included in Protein Pack were implemented as described by Cavanagh and co-workers (*26*) with minor modifications to include the use of PFGs for artifact suppression and coherence order selection of ^{15}N−^{1}H or ^{13}C−^{1}H magnetization with sensitivity enhancement using the approach of Rance (*27*, *28*) and Kay (*29*). This applied to the three-dimensional HCCH-COSY, HCCH-TOCSY, and HNHA experiments. The HNHB experiment was carried out according to the method of Archer and co-workers (*30*). Heteronuclear experiments in assigning aromatic resonances [(HB)CB(CGCD)HD and related experiments (Table 1)] were implemented on the basis of the original work of Yamazaki and co-workers (*31*) but were modified to include use of PFGs, as described above for other experiments. Measurements of three-bond ^{13}C−^{13}C and ^{15}N−^{13}C heteronuclear couplings were made using spin-echo difference constant-time (CT) HSQC experiments developed by Bax and co-workers (*32*). Further, three-bond ^{13}C−^{13}C couplings between ^{13}Cα and ^{13}Cδ for Leu residues were

Table 1: Acquisition Parameters for NMR Experiments

experiment	nucleus			no. of data points			spectral width (Hz)			no. of transients
	t1	t2	t3	t1	t2	t3	t1	t2	t3	
CBCA(CO)NH	^{13}C	^{15}N	^{1}H	32	64	1024	9000	1500	8000	8
HNCACB	^{13}C	^{15}N	^{1}H	32	64	1024	9000	1500	8000	16
C(CO)NH-TOCSY	^{13}C	^{15}N	^{1}H	32	64	1024	9500	1500	8000	16
H(CCO)NH-TOCSY	^{1}H	^{15}N	^{1}H	32	64	1024	6000	1500	8000	16
HNCO	^{13}CO	^{15}N	^{1}H	32	64	1024	1600	1500	8000	8
HN(CA)CO	^{13}CO	^{15}N	^{1}H	32	32	1024	1600	1500	8000	64
HCCH-TOCSY	^{1}H	^{13}C	^{1}H	128	50	1024	6000	6000	8000	8
HCCH-COSY	^{1}H	^{13}C	^{1}H	96	48	1024	6000	6000	8000	8
^{15}N-edited NOESY-HSQC	^{1}H	^{15}N	^{1}H	128	32	1024	6000	1550	8000	8
^{13}C-edited NOESY-HSQC	^{1}H	^{13}C	^{1}H	128	48	1024	6000	6000	8000	8
HNHA	^{1}H	^{15}N	^{1}H	32	64	1024	1500	6000	8000	8
HNHB	^{1}H	^{15}N	^{1}H	32	64	1024	1500	6000	8000	8
(HB)CB(CGCD)HD	^{13}C	^{1}H		16	1024		2000	6000		512
HB(CBCGCD)HD	^{1}H	^{1}H		16	1024		1500	6000		512
(HBCBCG)CDHD	^{13}C	^{1}H		16	1024		1500	6000		512
aromatic ^{13}C-edited NOESY-HSQC	^{1}H	^{13}C	^{1}H	256	64	1024	6000	6000	8000	8
^{13}C-{^{15}N}-CT-^{13}C-HSQC	^{13}C	^{1}H		256	1024		5000	6000		512
^{15}N-{^{13}C}-CT-^{13}C-HSQC	^{13}C	^{1}H		256	1024		5000	6000		512
3D LRCH	^{1}H	^{13}C	^{1}H	256	256	512	8000	8000	8000	16

measured using a three-dimensional ^{1}H-detected long-range correlation experiment (*32*). Quadrature detection in the indirectly detected dimensions was achieved either using the States−TPPI method (*33*) or using echo−antiecho selection with PFGs, as described above. Finally, PFG diffusion experiments were performed using a pulse sequence developed by Wu and co-workers (*34*) that incorporates bipolar gradient pulses to minimize eddy current effects. The diffusion coefficient was obtained by monitoring the signal amplitude as a function of gradient strength and fitting these data to the equation $S(q) = S(0)\exp[-Dq^2(\Delta - \delta/3 - \tau/2)]$ (*34*), where $S(q)$ is the NMR signal amplitude with a nonzero gradient strength, $S(0)$ is the NMR signal amplitude with a zero gradient strength, D is the diffusion coefficient, q is the gradient pulse area given as $q = \gamma g\delta$ (where γ is the gyromagnetic ratio, g is the gradient strength, and δ is the gradient duration), and τ is the inter-pulse delay time. Gradient strengths were calibrated using the diffusion coefficient of 1% H_2O in D_2O at 25 °C (*35*). A list of the NMR experiments used in this study and the relevant acquisition parameters is given in Table 1.

Data Processing and Analysis. Two- and three-dimensional NMR spectra were processed using the program Felix 98 (Molecular Simulations, Inc., San Diego, CA). In general, time domain data were apodized using sine-bell or shifted sine-bell functions, followed by zero-filling, and Fourier transformation. Non-CT heteronuclear time-domain data were first extended using linear prediction, while CT data were extended using mirror-image linear prediction (*35*). An external standard of 10 mM TSP in 200 mM potassium phosphate at pH 6.5 and 28 °C was used to reference the ^{1}H chemical shift scale. ^{13}C and ^{15}N chemical shifts were indirectly referenced using the following ratios of gyromagnetic ratios: 0.101 329 00 for ^{15}N to ^{1}H and 0.251 449 54 for ^{13}C to ^{1}H (*26*).

Sequence-specific assignments of ^{1}HN, ^{15}NH, ^{13}Cα, and ^{13}Cβ nuclei were established through the analysis of several pairs of spectra, including three-dimensional CBCA(CO)-NH and CBCANH, three-dimensional HN(CO)CA and HNCA, and three-dimensional HNCO and HN(CA)CO spectra. Side chain assignments were made through the analysis of three-dimensional (H)C(CO)NH-TOCSY, H(C-CO)NH-TOCSY, HBHA(CBCACO)NH, HCCH-COSY, and HCCH-TOCSY spectra. Aromatic proton and carbon assignments were made on the basis of the correlation to ^{1}Hβ and ^{13}Cβ nuclei using the following two-dimensional spectra: (HB)CB(CGCD)HD, HB(CBCGCD)HD, and (HBCB-CG)CDHD.

Stereospecific assignments of valine methyl groups were made through the analysis of two spin-echo difference CT-^{13}C HSQC spectra that allow the three-bond coupling constants between the methyl carbons and either the backbone ^{15}N or ^{13}C to be estimated (*32*). Stereospecific assignments of Val methyl groups were made for the following residues: V4, V14, V37, V40, V48, V49, V55, and V62. Stereospecific assignments of methylene Hβ protons were obtained through the analysis of three-dimensional HNHB and three-dimensional NOESY-^{15}N-HSQC and three-dimensional NOESY-^{13}C-HSQC spectra. Hβ protons were stereospecifically assigned for the following residues: Y6, S11, N12, D13, L15, N16, K18, L19, K29, D30, D43, L44, N46, N50, N52, L56, S65, L69, D77, and N81. Stereospecific assignments of leucine methyl groups were determined using three-dimensional NOE data combined with the analysis of $^{3}J_{C\alpha C\delta}$ coupling constants obtained from the relative intensity of the Cα and Cδ cross-peaks in a three-dimensional long-range ^{13}C−^{13}C correlation spectrum (*36*). Stereospecific methyl assignments were made for L15, L44, L56, L61, L69, and L85.

Distance restraints were obtained through the analysis of heteronuclear three-dimensional NOESY spectra. These included three-dimensional ^{1}H−^{15}N and ^{1}H−^{13}C NOESY-HSQC spectra with a mixing time of 100 ms. In processing the three-dimensional NOESY data, linear prediction was used to extend the interferograms by 50% in the ^{1}H dimension and by 50% in the ^{13}C dimension or by 100% in the ^{15}N dimension.

Structure Calculation. Interproton distances were estimated from the volumes of cross-peaks in three-dimensional NOESY spectra acquired with a mixing time of 100 ms. The volumes of cross-peaks corresponding to known interproton distances in regions of regular secondary structure (in this

Table 2: Structural Statistics

total no. of NOEs	3186
intraresidue	942
interresidue	
sequential	704
medium-range	954
long-range	586
no. of dihedral restraints	
ψ	64
ϕ	64
no. of H bonds	76
rmsd from mean structure (Å)	
backbone	0.5110
backbone (residues 4−92)	0.3745
all heavy atoms	0.8992
all heavy atoms (residues 4−92)	0.7572
violations	
distance restraints violations	
no. greater than 0.25 Å	0
torsion angle violations	
ψ no. greater than 5°	0
ψ no. greater than 5°	0
maximum distance restraints violation	0.24
Xplor energies (ensemble) (kcal/mol)	
total energy	269.842 ± 1.850
bond length	17.013 ± 0.200
bond angle	87.750 ± 1.295
improper	6.199 ± 0.544
van der Waals	111.236 ± 1.186
NOE	46.688 ± 1.656
dihedral	0.961 ± 0.0627
Ramachandran statistics (from Procheck NMR) (%)	
most favored region	81.9
allowed region	15.7
generously allowed	2.1
disallowed	0.3

study, α-helices) were used to calibrate the volume versus distance relationship. The restraint lower bounds were set to the van der Waals distance (1.8 Å), and upper bounds were set to the estimated distance + 20%. Estimates of the ϕ and φ backbone dihedral angles were obtained using the program TALOS, developed and distributed by Cornilescu and co-workers (*37*), using the chemical shifts of ^{1}Hα, ^{15}NH, ^{13}Cα, ^{13}Cβ, and ^{13}C′. TALOS-based ϕ and φ restraints were used only when a well-defined, consistent prediction was obtained. In practice, ϕ and φ restraints were limited to regular secondary structures. In addition, H-bond restraints were used within α-helices and the β-sheet in the following way. Two distance restraints were used for each hydrogen bond: one between the hydrogen and the acceptor atom (1.5−2.3 Å) and one between the donor heavy atom and the acceptor atom (2.4−3.3 Å). These were introduced after the initial set of structure calculations. Amide protons involved in H-bonds were identified on the basis of slow exchange with D_2O.

A total of 3186 NOE interproton distance restraints, 64 pairs of ϕ and φ backbone dihedral angle restraints, and 38 pairs of H-bond restraints were obtained (Table 2). MotNF structures were calculated from these restraints using X-PLOR (*38*) and torsion angle dynamics (TAD) (*39*). A total of 200 structures were calculated using MSI's version of X-PLOR, XPLOR 98. A modified version of the original TAD protocol (*39*) was used as follows: 15 ps high-temperature TAD (50 000 K) followed by cooling to 1000 K over the course of 15 ps and ramping of the van der Waals scaling term from 0.1 to 1.0. The molecules were further cooled to 300 K over the course of 6 ps using conventional Cartesian dynamics followed finally by 1000 steps of conjugate gradient energy minimization. The NOE energy term was 150 kcal/mol for the first three steps and 100 kcal/mol for the last. Several rounds of calculations were performed followed by analysis of violations and adjustment of restraints. Two hundred structures were calculated during each round. Statistics for the 20 lowest-energy structures from the final round of calculations are given in Table 2.

RESULTS

Oligomeric State of the MotNF Domain in Solution. Gel filtration chromatography revealed that the molecular mass of MotNF in solution is approximately 12 450 Da (Figure 1). This is in reasonable agreement with the calculated molecular mass of 10 234 Da based on the amino acid sequence, and suggests that MotNF exists as a monomer in solution. The disparity between the estimated and calculated molecular masses is probably due to the nonspherical shape of MotNF that gives rise to somewhat nonideal hydrodynamical characteristics. The molecular mass was also independently analyzed by dynamic light scattering and NMR. Light scattering yielded a calculated hydrodynamic radius of 1.8 nm and a corresponding molecular mass of 12 000 Da. NMR yielded a diffusion coefficient of 11.6×10^{-7} cm^2/s which is slightly larger than that for lysozyme (10.4×10^{-7} cm^2/s) that has a mass of 14 100 Da. This indicates that MotNF diffuses more rapidly than lysozyme in solution and is therefore a slightly smaller particle based on hydrodynamics. These two results are in agreement with the gel filtration results and confirm that MotNF is a monomer in solution under the conditions used for NMR spectroscopy and structure determination.

Assignments and Secondary Structure. The ^{15}N HSQC spectrum of MotNF (Figure 2) shows well-dispersed and sharp resonances, indicating that the protein is well-folded and monodisperse. Virtually complete sequence-specific backbone and side chain assignments (96% for ^{1}H, 98% for ^{15}N, and 98% for ^{13}C) were obtained by extensive analysis of a series of multidimensional heteronuclear NMR spectra (Table 1) and are reported elsewhere (N. Li et al., unpublished results, 2000). These assignments allowed near-complete analysis of the resonances in three-dimensional NOESY spectra. The patterns of short- and medium-range NOEs observed for MotNF are shown in Figure 3 and clearly reveal five α-helices. The deviations of ^{13}Cα chemical shifts from random-coil values (termed the secondary chemical shift, $\Delta\delta$ ^{13}Cα) are reliable indicators of secondary structure (*40*); for α-helices, $\Delta\delta$ ^{13}Cα values are between 2 and 6 ppm and for β-strands are between −1 to −3 ppm. The $\Delta\delta$ ^{13}Cα values for MotNF (Figure 4A) identify the same five α-helices discussed above on the basis of NOEs and, in addition, clearly identify two short β-strands (Figure 4A).

Structure of MotNF. A superposition of the 20 lowest-energy structures is shown in Figure 5A. A composite Ramachandran plot for these structures indicates that 99.7% of the backbone dihedral angles are in allowed or generously allowed regions, and 0.3% in the disallowed regions according to the program PROCHECK-NMR (*41*). The structures are well-defined and show excellent agreement with the NOE data; there are no violations of distance restraints greater than 0.25 Å (Table 2). The coordinate rmsd for backbone heavy

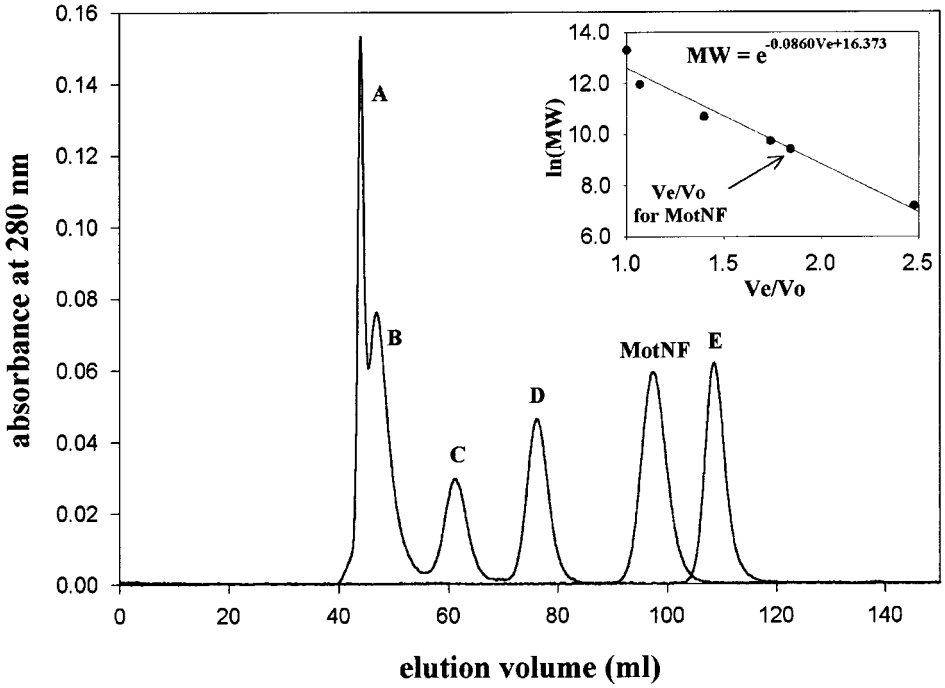

FIGURE 1: Gel filtration analysis of the oligomeric state of the MotNF domain using a Sephadex 75 column (Pharmacia). Two chromatographic traces are overlaid, one for the molecular mass standards (Bio-Rad) and another for MotNF. The peaks corresponding to the molecular mass standards are labeled A−E, and the peak for MotNF is also labeled. The approximate molecular masses of the reference proteins are as follows: (A) 600 000, (B) 158 000, (C) 44 000, (D) 17 000, and (E) 1350 Da. The inset shows the linear regression of the data from peaks A−E.

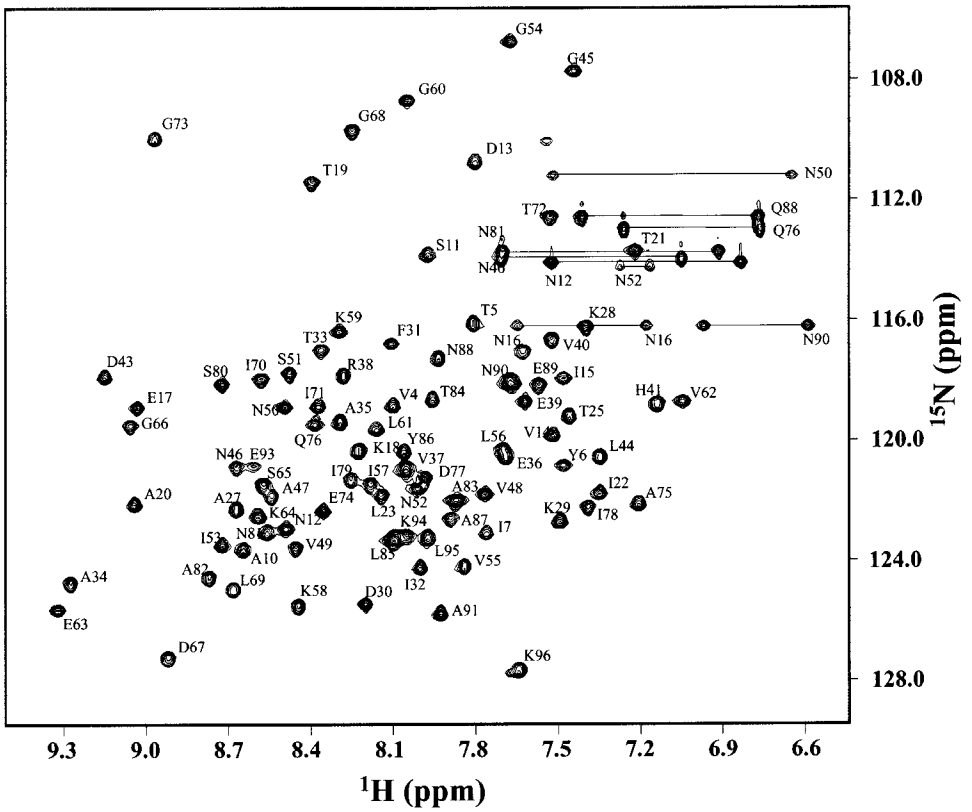

FIGURE 2: ^{15}N-labeled MotNF domain HSQC spectrum. Peaks connected by lines are from the amide protons of side chain Asn and Gln.

atoms of the folded core of MotNF (residues 4−92) is 0.37 Å. An average of 33 unique restraints per residue was used in the final structure calculations, and this factor accounts for the high precision of the structures. However, since the restraints are not evenly distributed throughout the molecule (Figure 4B), the regions of the protein that have a larger number of restraints per residue are more well defined (Figure 4C). Residues at the extreme C-terminus (91−96) exhibit a higher than average rmsd between structures due to a general lack of long-range restraints in this region (Figure 4B,C). The C-terminal helix is broken by a Pro residue at position 92, and the final six residues probably form the

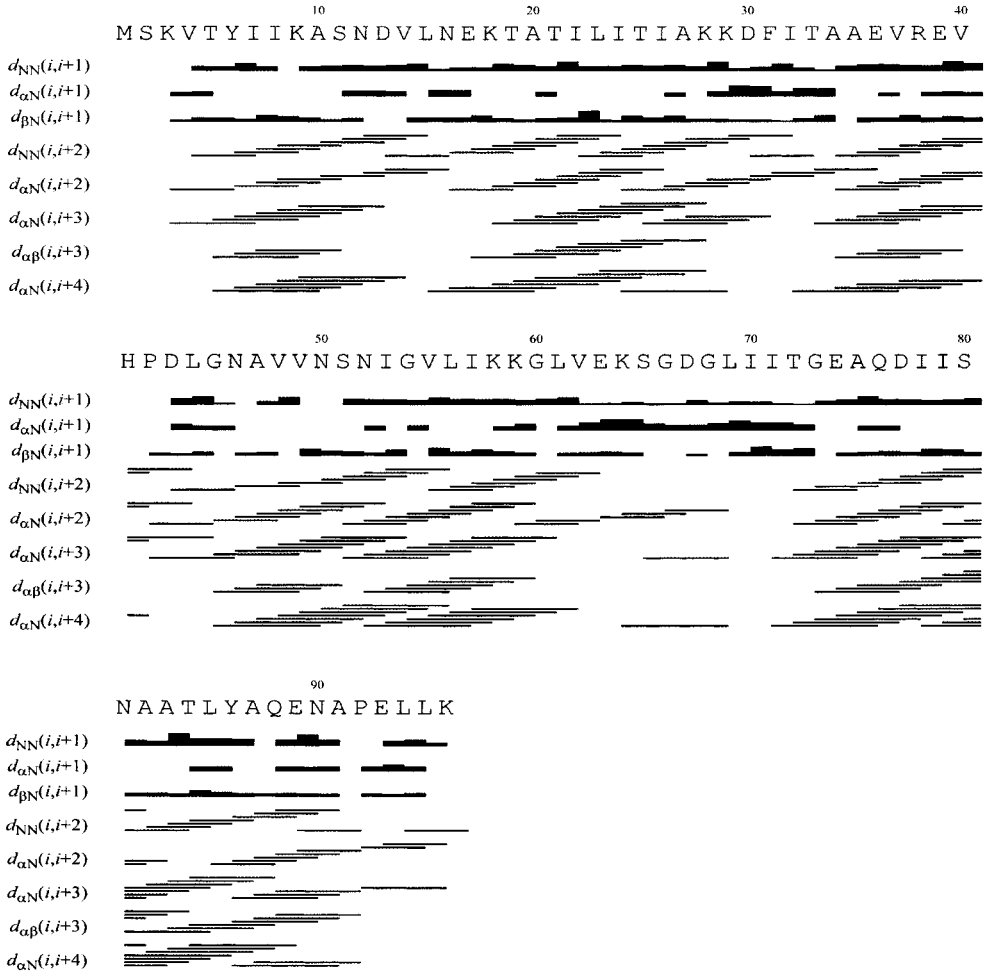

FIGURE 3: Schematic representation of the sequential and medium-range NOE connectivities involving NH, Hα, and Hβ protons for the MotNF domain. The thickness of the bar indicates the intensity of NOEs.

beginning of a linker between the N- and C-terminal domains of MotA and may be flexible in the context of the full protein.

It is important to emphasize that the known crystal structure of MotNF (*2*) was never consulted during the NMR analysis. It is therefore gratifying that the overall structure of the MotNF domain in solution is very similar to the X-ray structure (Figure 5B). The rmsd between the two structures is 2.17 Å for the backbone atoms, and 3.72 Å for all heavy atoms. Briefly, MotNF consists of a bundle of five α-helices in which four [α1(4−11), α3(34−40), α4(46−59), and α5-(73−90)] are amphipathic and pack their hydrophobic surfaces around the central helix, α2(17−28) (Figure 5B). There are two regions where there are significant differences between the solution and crystal structures. One is at the C-terminus, and the other is within the loop region encompassing residues 40−47. In the NMR structure, helix α5 ends at residue 90 and is followed by six unstructured terminal residues. This is in contrast to the crystal structure where helix α5 ends at residue 91, and the terminal four residues form the single-turn helix α6 separated from α5 by Pro92 (Figure 5B, left vs right panel). With regard to the loop region, the differences between the average NMR structure and the X-ray structure may be due to the smaller number of restraints per residue that are available for this region which results in significantly higher rmsd values within the ensemble of 20 NMR structures. When these two regions are excluded, the rmsd of the core structures between the NMR and crystal structure is 1.12 Å for the backbone atoms and 3.08 Å for all heavy atoms. Finally, it should also be noted that the angle between helix α2 and helix α5 is 80° in the NMR structure which differs slightly from the value of 85° observed in the crystal structure.

DISCUSSION

Our original assumption that the MotNF crystallographic structure is physiologically important was based on several attractive reasons for why full-length MotA should bind cognate DNA as a dimer. It has been established that MotA forms multimers when bound to DNA, and a dimer would be a logical first step in this process (*9*, *42*). Also, dimerization upon binding DNA is quite common in transcription factors since it effectively increases both the specificity and the association constant of the interaction (*22*, *23*). Finally, our preliminary NMR analysis of the structure of the C-terminal DNA-binding domain of MotA suggests that it resembles one-half of the eukaryotic TATA-binding protein (TBP), and dimerization could generate a TBP-like domain (*20*). This is appealing because middle-mode transcription (*11*), and several other T4 processes (*43*−*45*), appear to function more like eukaryotic rather than prokaryotic systems.

However, a recent detailed analysis of the DNA-binding properties of MotA using a variety of independent techniques

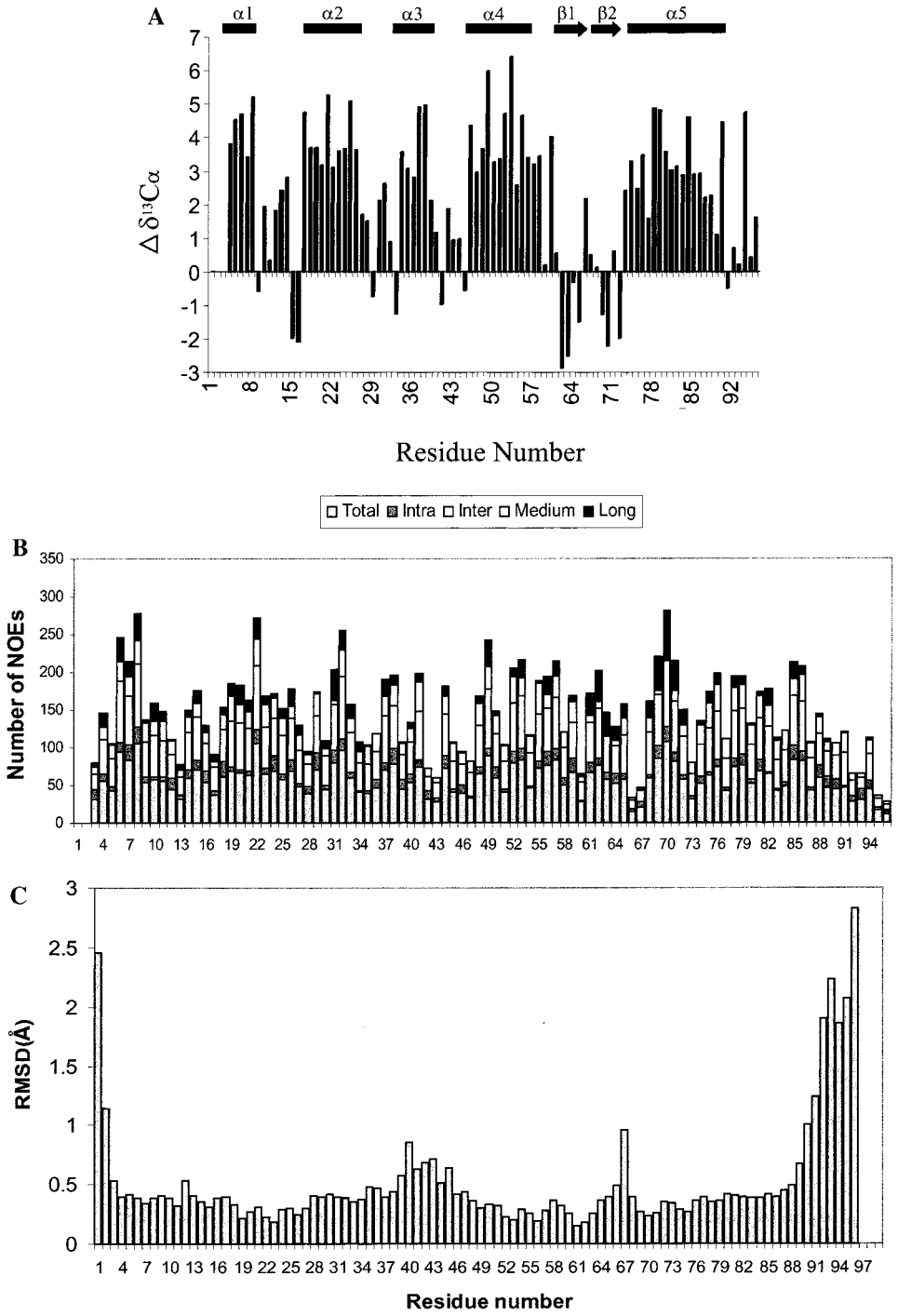

FIGURE 4: NMR data for each of the amino acids in the MotNF domain. (A) $^{13}C\alpha$ secondary shift [difference between the observed Cα chemical shift and the published random-coil value (*40*)] for each residue. (B) Number of unique (unambiguous) restraints plotted vs residue number. (C) The rmsd per residue for the family of 20 NMR structures shown in Figure 5A.

clearly shows that a single MotA molecule binds to the mot box (*24*). This study also confirmed that, while MotA does form multimers on an extended DNA segment that contains the mot box, these multimers are nonspecific, weakly bound, and not dependent on whether the additional DNA is upstream or downstream of the mot box. This latter finding is crucial because the rigid MotNF dimer found in the crystal demands that a second MotA binding site be positioned on a particular side of the Mot box, and this is not the case. Further, recent studies have shown that the molecular bases for recognition of DNA by MotA and TBP are different (*12*), weakening the hypothesis that the two proteins function similarly as dimers in binding DNA. Finally, one would predict from the intimate coiled-coil interaction between the monomers in the crystal structure that multimer formation would be a cooperative process, but this was shown not to be the case. Subsequent studies on the mot box sequence have confirmed that it is restricted to nine base pairs (*17*), and one would expect this consensus sequence to be more extensive if two MotA proteins were involved in the binding. These recently published data convinced us of the importance of performing these structural studies of MotNF in solution.

Our results confirm that MotNF is indeed a monomer in solution, and they also suggest two reasons why the domain

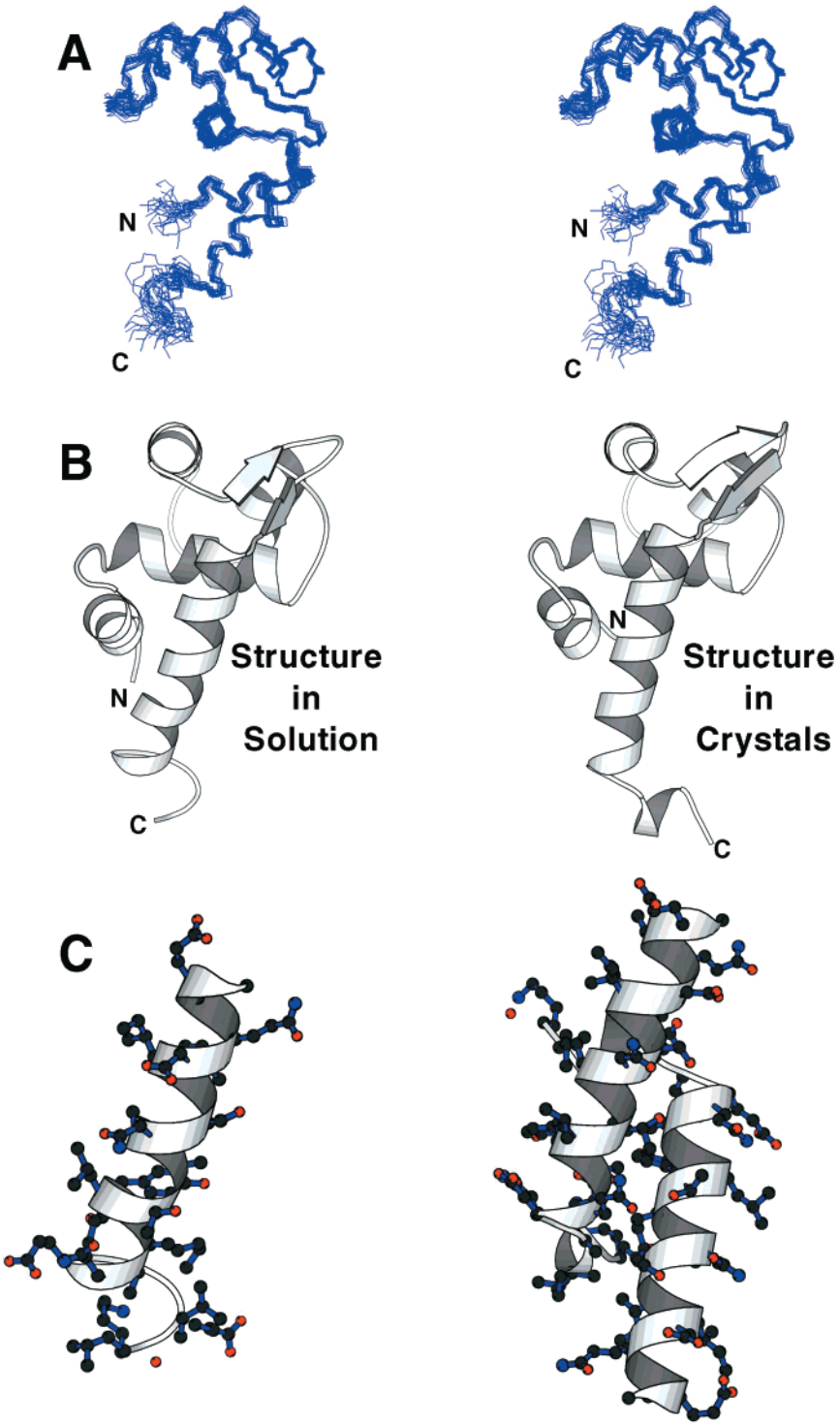

FIGURE 5: Structure of the MotNF domain. (A) Stereoview of the 20 lowest-energy structures for MotNF in solution. (B) Ribbon diagram of the average of 20 structures shown in panel A (left) and ribbon diagram of the crystal structure of the MotNF domain subunit (right). (C) C-Terminal α-helix of the NMR solution structure (left) and C-terminal α-helices in the crystal structure (right).

forms a biologically compelling dimer in the crystals. First, helix α5 is ideally suited for the formation of an antiparallel coiled-coil structure between two MotNF molecules. It extends away from the core and appears to have some flexibility in its orientation, and one surface contains an array of hydrophobic residues that promotes a reciprocal coiled-coil interaction (Figure 5C). Presumably, the actual role of helix α5 in the full-length MotA protein is to act as a rigid connection between the two domains. The second reason is that the crystal dimer stabilizes one turn of α-helix at the extreme C-terminus following Pro92 (α6 in the crystal structure), and this also contributes to the dimer interface (Figure 5C). In the context of the complete MotA protein, the residues within α6 may actually constitute a linker region between the domains, or simply be a portion of the C-terminal domain. The crystal structure of the C-terminal half of MotA, for which we now have crystals (unpublished data), may help to deduce the true conformation of these amino acids. This scenario for dimer formation is supported by the comparison of the NMR and crystal structures of

MotNF where the cores are very similar, and the crucial differences are localized to the C-terminal region (Figure 5B). Importantly, none of the C-terminal interfacial structural features are observed in solution on the basis of our NMR data.

The confirmation that MotA functions as a monomer is consistent with recent insights into its functional role at T4 middle promoters. The domain structure of σ^{70} has now been established (*46*), and MotA interacts with region 4.2 that normally binds to the −35 region of *E. coli* promoters (*12*, *47*). The binding of AsiA at, or close to, region 4.2 of σ^{70} prevents this interaction, and MotA acts as an adapter protein that effectively converts the specificity of σ^{70} to the mot box. Thus, MotA is not acting as a transcription factor per se, but more as a transcription-associated factor. Our original proposal was that the function of MotA might resemble eukaryotic transcription factors based on the similarity of its DNA-binding domain and the TATA-binding protein (*20*). This now seems unlikely, but it has been pointed out that the MotA/AsiA system can now be considered as a prokaryotic version of the TAF-mediated promoter switching process that is employed by eukaryotic polymerases (*17*). Thus, a relationship between T4 middle-mode transcription and eukaryotic transcription is still a possibility.

Subdividing proteins into independently folded domains for structural analysis is a widely used technique, but our results have shown that this can lead to artifacts with functional consequences. This problem is particularly relevant to the emerging structural genomics effort where structure, and not function, is the emphasis. In the case presented here, conclusions based on observing the intermolecular dimer in crystals led to erroneous conclusions about the role of dimerization in DNA binding. The artifactual dimer interface is only generated as a result of removing the C-terminal domain, and does not exist in the full-length protein. This study highlights how crystal packing can influence protein conformation, and emphasizes the need to examine protein structures in both crystals and solution to fully understand biomolecular structure and to accurately deduce structure−function relationships. No matter how compelling, our experience with MotNF urges caution when interpreting intermolecular interactions observed in crystals, and demonstrates the power of a combined approach using both X-ray diffraction and NMR spectroscopy. Furthermore, careful analysis of the state of aggregation of a protein in solution can help resolve questions generated by structural data from crystals. In this study, we used gel filtration chromatography, dynamic light scattering, and NMR-based diffusion measurements to establish the molecular size of MotNF, and to rule out the existence of the dimer configuration in solution.

ACKNOWLEDGMENT

We thank members of the Kriwacki and White laboratories for assistance, comments, and criticism during the course of this study and the support staff at MSI, including Drs. Steve Unger, Jian Xu, and Tina Yeh, for assistance with Felix97 and XPLOR98 software.

REFERENCES

1. Schafer, P. H., Pierce, S. K., and Jardetzky, T. S. (1995) *Semin. Immunol. 7*, 389−398.
2. Finnin, M. S., Cicero, M. P., Davies, C., Porter, S. J., White, S. W., and Kreuzer, K. N. (1997) *EMBO J. 16*, 1992−2003.
3. Wilkens, K. P., and Ruger, W. (1994) in *Molecular Biology of Bacteriophage T4* (Karam, J. W., Ed.) pp 132−141, ASM Press, Washington, DC.
4. Williams, K. P., Kassavetis, G. A., Herendeen, D. R., and Geiduschek, E. P. (1994) in *Molecular Biology of Bacteriophage T4* (Karam, J. W., Ed.) pp 161−175, ASM Press, Washington, DC.
5. Stitt, B., and Hinton, D. (1994) in *Molecular Biology of Bacteriophage T4* (Karam, J. W., Ed.) pp 142−160, ASM Press, Washington, DC.
6. deFranciscis, V., and Brody, E. (1982) *J. Biol. Chem. 257*, 4087−4096.
7. deFranciscis, V., Favre, R., Uzan, M., Leautey, J., and Brody, E. (1982) *J. Biol. Chem. 257*, 4097−4101.
8. Hinton, D. M. (1991) *J. Biol. Chem. 266*, 18034−18044.
9. Schmidt, R. P., and Kreuzer, K. N. (1992) *J. Biol. Chem. 267*, 11399−11407.
10. Orsini, G., Ouhammouch, M., Le Caer, J.-P., and Brody, E. N. (1993) *J. Bacteriol. 175*, 85−93.
11. Ouhammouch, M., Adelman, K., Harvey, S. R., Orsini, G., and Brody, E. N. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 1451−1455.
12. Sharma, M., Marshall, P., and Hinton, D. M. (1999) *J. Mol. Biol. 290*, 905−915.
13. Stevens, A. (1976) in *RNA Polymerase* (Losick, R., and Chamberlin, M., Eds.) pp 617−627, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
14. Stevens, A. (1977) *Biochim. Biophys. Acta 475*, 193−196.
15. Colland, F., Orsini, G., Brody, E. N., Buc, H., and Kolb, A. (1998) *Mol. Microbiol. 27*, 819−829.
16. Uzan, M., d'Aubenton-Carafa, Y., Favre, R., deFranciscis, V., and Brody, E. (1985) *J. Biol. Chem. 260*, 633−639.
17. Marshall, P., Sharma, M., and Hinton, D. M. (1999) *J. Mol. Biol. 285*, 931−944.
18. Kolesky, S., Ouhammouch, M., Brody, E. N., and Geiduschek, E. P. (1999) *J. Mol. Biol. 291*, 267−281.
19. Finnin, M. S., Hoffman, D. W., Kreuzer, K. N., Porter, S. J., Schmidt, R. P., and White, S. W. (1993) *J. Mol. Biol. 232*, 301−304.
20. Finnin, M. S., Hoffman, D. W., and White, S. W. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10972−10976.
21. Gerber, J. S., and Hinton, D. M. (1996) *J. Bacteriol. 178*, 6133−6139.
22. Ptashne, M., and Gann, A. A. (1990) *Nature 346*, 329−331.
23. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
24. Cicero, M. P., Alexander, K. A., and Kreuzer, K. N. (1998) *Biochemistry 37*, 4977−4984.
25. Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) *J. Bacteriol. 119*, 736−747.
26. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy*, Academic Press, New York.
27. Cavanagh, J., and Rance, M. (1990) *J. Magn. Reson. 88*, 72−85.
28. Palmer, A. G. I., Cavanagh, J., Wright, P. E., and Rance, M. (1991) *J. Magn. Reson. 93*, 151−170.
29. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
30. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) *J. Magn. Reson. 95*, 636−641.
31. Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 11054−11055.
32. Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R., and Zhu, G. (1994) *Methods Enzymol. 239*, 79−105.
33. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
34. Wu, D., Chen, A., and Johnson, C. S. J. (1995) *J. Magn. Reson., Ser. A 115*, 260−264.

35. Hoch, J. C., and Stern, A. S. (1996) *NMR Data Processing*, Wiley-Liss, New York.
36. Bax, A., Max, D., and Zax, D. (1992) *J. Am. Chem. Soc. 114*, 6923−6925.
37. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR 13*, 289−302.
38. Brünger, A. T. (1992) *X-PLOR*, Yale University Press, New Haven, CT.
39. Stein, E. G., Rice, L. M., and Brünger, A. T. (1997) *J. Magn. Reson. 124*, 154−164.
40. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
41. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
42. March-Amegadzie, R., and Hinton, D. M. (1995) *Mol. Microbiol. 15*, 649−660.
43. Spicer, E. K., Rush, J., Fung, C., Reha-Krantz, L. J., Karam, J. D., and Konigsberg, W. H. (1988) *J. Biol. Chem. 263*, 7478−7486.
44. Tsurimoto, T., and Stillman, B. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1023−1027.
45. Huff, A. C., and Kreuzer, K. N. (1991) in *DNA Topoisomerases in Cancer* (Potmesil, M., and Ross, W., Eds.) Oxford University Press, New York.
46. Severinova, E., Severinov, K., Fenyo, D., Marr, M., Brody, E. N., Roberts, J. W., Chait, B. T., and Darst, S. A. (1996) *J. Mol. Biol. 263*, 637−647.
47. Hochschild, A., and Dove, S. L. (1998) *Cell 92*, 597−600.

BI0028284