Dynamics of the Mrf-2 DNA-Binding Domain Free and in Complex with DNA†

Lingyang Zhu,‡ Jingui Hu,‡,§ Donghai Lin,‡ Robert Whitson,|| Keiichi Itakura,|| and Yuan Chen*,‡

Division of Immunology and Division of Biology, Beckman Institute of the City of Hope, Duarte, California 91010

Received March 8, 2001; Revised Manuscript Received May 30, 2001

ABSTRACT: Mrf-2 is a member of a new class of DNA-binding proteins known as the AT-rich interaction domain family or ARID. Chemical shift indices and characteristic NOE values indicate that the three-dimensional structure of the Mrf-2 ARID in complex with DNA is nearly identical to that of the free protein. The backbone dynamics of the Mrf-2 domain free and in complex with DNA have been characterized by ^{15}N NMR relaxation measurements and model-free analysis. Chemical shift perturbations and dynamic studies suggest that two flexible interhelical loops, the flexible C-terminal tail, and one α-helix are involved in DNA recognition, indicating the importance of protein dynamics in DNA binding. Some well-structured regions, in particular the putative DNA-contacting helix, in Mrf-2 show a decrease in the order parameters (S^2) upon complex formation. The less well-structured loops and the unstructured C-terminus show reduced flexibility upon DNA binding. In addition, the model-free analysis indicates motions on the picosecond to nanosecond and micro- to millisecond time scales at the DNA-binding surface of the bound Mrf-2 ARID, suggesting a model where interactions between the protein and DNA are highly dynamic.

Protein−DNA interactions are important to many biological processes such as transcriptional regulation, DNA replication, and repair. Although three-dimensional structures of proteins, DNA, and their complexes provide significant insights into the determinants of binding affinity and specificity, dynamics clearly play important roles in molecular recognition (*1*, *2*). In many systems of protein−DNA interactions and other molecular complexes, the binding interfaces often involve flexible loops and termini. These flexible regions undergo local folding that is coupled to complex formation. The local folding transitions are suggested to be important to binding specificity, as nonspecific interactions may not be able to generate such "induced fits". In addition, changes in flexibility may influence the affinity of the interaction through changes in entropy and its contribution to free energy changes. In many cases, regular secondary structures appear to be important for the overall architecture of a protein, but flexible regions may be more important for interactions with other molecules and function.

Mrf-2 is a member of a recently identified family of DNA-binding proteins that contain ARID[1] (AT-rich interaction domains). This family includes more than 30 proteins from yeast, *Arabidopsis*, *C. elegans*, *Drosophila*, mice, and humans (*3*−*5*). Mrf-2 specifically recognizes the modulator region of the major immediate-early gene (MIE) of human cytomegalovirus (HCMV). The solution structure of the Mrf-2 domain contains six helices, two loops, and a flexible C-terminus (*6*). In the DNA complex, Mrf-2 contacts DNA in both the major and minor grooves (*7*). Although Mrf-2, Bright, and Dead ringer recognize specific DNA sequences, this has not been demonstrated for other members of the ARID family. The SWI1 ARID shows no sequence specificity (*8*).

Characterization of the interactions between the Mrf-2 ARID and target DNA is necessary for a detailed understanding of the manner in which this family of proteins recognizes DNA sequences. This paper describes the chemical shift assignments and structural characterization of the Mrf-2 ARID in complex with DNA. It also summarizes dynamic studies of the free and the DNA-bound Mrf-2 as investigated by ^{15}N relaxation measurements. Analysis of changes in both chemical shifts and backbone dynamics indicates that the DNA-binding sites on the Mrf-2 ARID mainly involve flexible protein regions including two loops and the C-terminus. In addition, an α-helix next to one of the flexible loops also appears to be involved in DNA recognition.

MATERIALS AND METHODS

Sample Preparation. ^{15}N-labeled Mrf-2 and the triply labeled ^{15}N/^{13}C/^{2}H Mrf-2 were expressed and purified as described previously (*6*). The modified plasmid has an open

† This work is supported by NIH Grants GM54190 and GM59887 (Y.C.). Y.C. is a recipient of the American Cancer Society Junior Faculty Research Award.

* To whom correspondence should be addressed. E-mail: ychen@coh.org. Fax: 626-301-8186. Phone: 626-930-5408.

‡ Division of Immunology.

§ Current address: Material Research Laboratory, University of California, Santa Barbara, CA 93106.

|| Division of Biology.

[1] Abbreviations: NMR, nuclear magnetic resonance; R_1, longitudinal relaxation rate constant; R_2, spin−spin relaxation rate constant; NOE, nuclear Overhauser effect; Mrf-2, modulator recognition factor 2; ARID, AT-rich interaction domain; HSQC, heteronuclear single quantum coherence; NOESY, nuclear Overhauser enhancement spectroscopy; ppm, parts per million; DTT, dithiothreitol; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate; CSA, chemical shift anisotropy; DD, dipole−dipole interaction.

10.1021/bi010476a CCC: $20.00
Published on Web 07/12/2001

reading frame that encodes the sequence Met-Arg-Gly-Ser-$(His)_6$-Gly-Ser, followed by residues 17−123 of Mrf-2 (GenBank accession number AAA59870). The residues are numbered from the N-terminus of the expressed peptide. The isotopically enriched proteins were expressed in appropriate M9 minimal media, yielding 100% ^{15}N labeling or 100% $^{15}N/^{13}C$ labeling with 70% deuteration. The DNA in the complex contained 16 base pairs with the sequence 5′-GCACAATATAACGTCG-3′ and its complement. The DNA was synthesized by standard solid phase methods and purified by HPLC at the DNA synthesis facility of the City of Hope. The NMR samples contained approximately 1 mM free Mrf-2 (^{15}N labeled only) or 0.6 mM complex with DNA ($^{15}N/^{13}C/^{2}H$-enriched Mrf-2 for assignment, and ^{15}N-enriched Mrf-2 for relaxation measurement) in 5 mM DTT, 5% D_2O/95% H_2O at pH 6.0. The free Mrf-2 sample also contained 100 mM sodium phosphate as buffer.

NMR Spectroscopy. NMR spectroscopy was performed on a Varian Unity-plus 500 MHz spectrometer equipped with a triple-resonance probe, pulse shaping, and pulsed-field gradient capabilities. The relaxation measurements on the free protein were conducted at 26 °C, and relaxation data of the complex were collected at 30 °C to obtain optimal data quality. DSS was used to indirectly reference the ^{15}N and ^{13}C resonances (*9*). NMR data were analyzed using Felix 97 (MSI) software. Resonance assignments of the complex were achieved from the three-dimensional (3D) spectra HNCOCA (*10*), HNCA, HNCACB (*11*), and ^{15}N edited NOESY−HSQC (*12*) with a 100 ms mixing time.

NMR Relaxation Measurements. The ^{15}N relaxation rate constants R_1 and R_2 and NOE experiments were measured using 2D inversion−recovery, CPMG, and steady-state NOE pulse sequences as described (*13*). A recycle delay of 2 s was used in both R_1 and R_2 measurements for Mrf-2 free and in complex with DNA. The relaxation delays in the R_1 experiments were 355 (×2), 710, 1070, 1420, and 1660 ms with a duplicate point at 355 ms (as indicated by ×2) for error estimation of the free Mrf-2 data, and were 11.1, 222, 444, 666, 888 (×2), 1110, 1332, and 1665 ms for the bound Mrf-2. The relaxation delays in the R_2 experiments were 16.7, 33.4 (×2), 50.2, 83.5, 116.9, and 150.3 ms for the free Mrf-2 and were 16.8, 33.6, 50.4, 67.2 (×2), 84.0, and 117.6 ms for the bound Mrf-2. The nitrogen 90° pulse was 75 μs, to reduce sample heating during the experiments. For the measurements of NOEs, two pairs of NOE spectra were collected for the two forms of Mrf-2, respectively. A recycle time of 5 or 6 s was used for the free and bound Mrf-2, respectively. For the spectra with proton saturation, 3 s proton presaturation was used. The program Curvefit (Arthur G. Palmer, III) was used to extract relaxation rates R_1 and R_2 by fitting the data to a single-exponential decay function. NOE values were calculated as the ratios of the intensities with and without proton saturation. Errors in R_1 and R_2 were estimated using the repeated data points and were generated during the nonlinear least-squares fitting. Errors in NOE were determined from duplicate experiments.

The cross-correlation rates between ^{15}N CSA and $^{15}N-^{1}H$ dipolar relaxation (CSA/DD), η_N, were measured to identify residues that undergo motions on the μs−ms time scale. The spectral density at zero frequency that is free from exchange contribution can be estimated from η_N (*14*, *15*). The experiment was performed at 30 °C using the pulse sequence described in reference (*14*) and was repeated 4 times. For each experiment, a reference and 3 CSA/DD data sets were recorded in an interleaved fashion using cross-correlation delay (4Δ) values of 15, 20, and 25 ms. The intensity ratios of the ^{15}N CSA/DD cross-peaks to the reference cross-peaks have a linear dependence to the cross-correlation rates as $I^{cross}/I^{ref} = 4\Delta\eta_N$ under above cross-correlation delay values (*14*). The rates η_N were extracted by 1-parameter linear fitting and averaged over 4 repeated measurements. The errors were determined from the standard deviations of the values from the 4 repeated experiments. The rate η_N is related to the spectral density as

$$\eta_N = dc[4J^{cc}(0) + 3J(\omega_N)]P_2(\cos\theta)$$

where d and c are defined as in reference (*15*) and θ is the angle between the $^{15}N-^{1}H$ bond vector and the principal axis of the ^{15}N chemical shift tensor, assumed to be axially symmetric. In this study, the θ angle was assumed to be 20°. The $J^{cc}(0)$ extracted from η_N is free from the exchange effect, whereas $J(0)$ calculated from R_1, R_2, and NOE may include the exchange contribution (*16*). By comparison of the values of $J^{cc}(0)$ and $J(0)$, those residues that experience conformational exchange were identified.

Analysis of the Relaxation Data. The amide ^{15}N relaxation is dominated by the dipolar interaction with its directly attached ^{1}H and by chemical shift anisotropy (*17*). The relevant R_1, R_2, and NOE parameters for this nucleus can be expressed as functions of spectral density $J(\omega)$ at certain frequencies [(*13*) and references cited therein]. The spectral density function describes the reorientation of an N−H vector. For an N−H bond experiencing isotropic overall tumbling (ns time scale) and an independent local internal motion (ps time scale), $J(\omega)$ can be written in the form of Lipari and Szabo's model-free approach (*18*):

$$J(\omega) = \frac{S^2\tau_m}{1+\omega^2\tau_m{}^2} + \frac{(1-S^2)\tau}{1+\omega^2\tau_m{}^2} \qquad (1)$$

where τ_m is the correlation time of the overall rotation of the molecule, S^2 is the generalized order parameter describing the amplitudes, and τ_e is the effective correlation time describing the time scales of the internal motion of an N−H bond vector, and $\tau^{-1} = \tau_m{}^{-1} + \tau_e{}^{-1}$. An extended model-free formalism can be used if the internal motion is separated further into two independent motions of very different time scales (*19*). The two-time-scale spectral density function is written as

$$J(\omega) = \frac{S^2\tau_m}{1+\omega^2\tau_m{}^2} + \frac{(S_f{}^2 - S^2)\tau}{1+\omega^2\tau^2} \qquad (2)$$

where $S^2 = S_f{}^2S_s{}^2$.

An initial estimation of the overall tumbling correlation time (τ_m) was obtained from the average R_2/R_1 ratio of the N−H bond vectors of the selected residues which are independent of internal motions on the nanosecond time scale and of conformational exchange (R_{ex}). Accordingly, residues were chosen based on the two criteria of their relaxation data: (a) NOE ≥ 0.7 and (b) R_1 and R_2 ($1/T_1$

and $1/T_2$) satisfy (*20*)

$$\frac{\langle T_2\rangle - T_{2i}}{\langle T_2\rangle} - \frac{\langle T_1\rangle - T_{1i}}{\langle T_1\rangle} \leq 1.5\sigma \qquad (3)$$

where $\langle T_1\rangle$ and $\langle T_2\rangle$ are the averages, T_{1i} and T_{2i} are the T_1 and T_2 for residue i, respectively, and σ is the standard deviation of $(\langle T_2\rangle - T_{2i})/\langle T_2\rangle - (\langle T_1\rangle - T_{1i})/\langle T_1\rangle$.

The motional parameters were extracted from least-squares fitting of the relaxation data using the program Modelfree 3.1 (Arthur G. Palmer, III). During the model-free analysis, the model-free parameters were optimized by minimizing χ^2 given by

$$\chi^2 = \sum_i \frac{(R_{1i}^{e} - R_{1i}^{c})^2}{\sigma_{1i}^{2}} + \frac{(R_{2i}^{e} - R_{2i}^{c})^2}{\sigma_{2i}^{2}} + \frac{(\mathrm{NOE}_i^{e} - \mathrm{NOE}_i^{c})^2}{\sigma_{\mathrm{NOE}i}^{2}} \qquad (4)$$

where R_{1i}^{e}, R_{2i}^{e}, and NOE_i^{e} are the experimental values of the relaxation parameters; R_{1i}^{c}, R_{2i}^{c}, and NOE_i^{c} are the corresponding calculated values; and σ_{1i}, σ_{2i}, and $\sigma_{\mathrm{NOE}i}$ are the corresponding experimental uncertainties in the relaxation data. The quality of the fit between the experimental data and theoretical models was assessed by comparing the optimal values of χ^2 with the 95% critical value of the distribution of χ^2 determined from Monte Carlo simulations (*21*). During this step, five motional "models" for each residue were evaluated, and one was selected using F statistical testing. The five models include: (1) S^2 only; (2) S^2 and τ_e; (3) S^2 and R_{ex}; (4) S^2, τ_e, and R_{ex}; and (5) S^2, S_f^2, and τ_e. The conformational exchange R_{ex} term describes microsecond to millisecond time-scale motions which cause line-broadening in NMR spectra. Finally, following the determination of the appropriate model for each N−H bond vector, the overall correlation time was optimized for all of the residues simultaneously. Meanwhile, the motional parameters were extracted for each residue by fitting their relaxation data into the selected models.

RESULTS

Resonance Assignments of the Mrf-2 Domain in Complex with DNA. Nearly all (92%) backbone HN, C_α, and N resonances and 80% C_β have been assigned using a combination of triple resonance experiments and an ^{15}N-resolved NOESY spectrum. The spectrum for R_1 measurement at a relaxation delay of 222 ms is shown in Figure 1 as a representative of the ^{1}H−^{15}N HSQC spectrum of the Mrf-2 ARID in complex with DNA. The secondary structure of the bound Mrf-2 ARID has been determined using C_α chemical shift indices and characteristic NOE values (*9*, *22*). The secondary structure of the Mrf-2 ARID when complexed with DNA is identical to that of the free protein, which contains six α-helices (H1−H6) and a kink between H5 and H6. The two loops (L1 and L2) and a flexible C-terminal tail observed in the free protein remain in the complex. No global chemical shift changes are observed in Mrf-2 upon complex formation with DNA. These results suggest that the Mrf-2 ARID in complex with DNA has a tertiary structure that is nearly identical to that of the free protein.

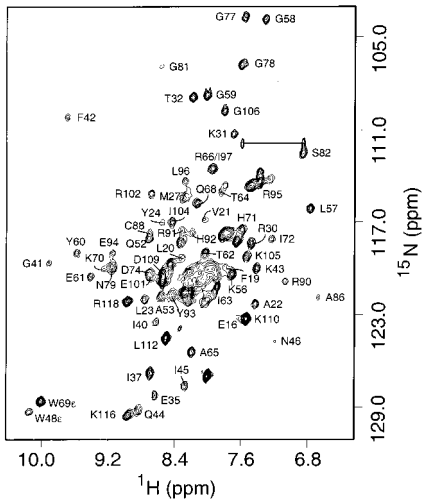

FIGURE 1: ^{15}N−^{1}H correlation spectrum of Mrf-2 in complex with DNA. Only the well-resolved cross-peaks, except that of the histidine-tag residues, are labeled with the assignments. This is one of the spectra used to calculate the relaxation rate constant R_1. The relaxation delay of this spectrum is 222 ms. All experiments for the complex were performed at 30 °C with 0.6 mM protein−DNA complex at pH 6.0. The complex consisted of ^{15}N/^{13}C/^{2}H-labeled Mrf-2 and equilmolar amounts of unlabeled 16 bp DNA (5′-GCACAATATAACGTCG-3′ and its complement).

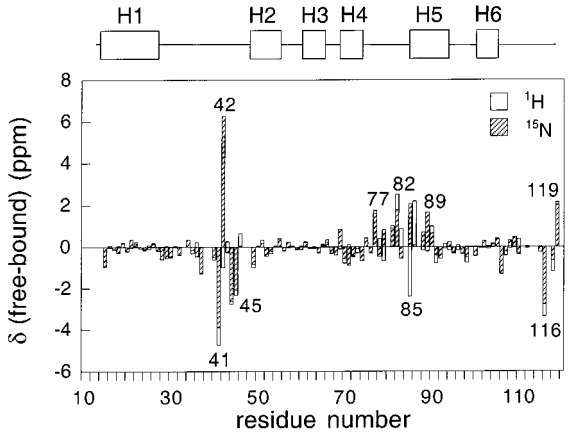

FIGURE 2: Chemical shift changes of amide ^{1}H (open bars) and ^{15}N (hatched bars) in Mrf-2 upon DNA binding. Residues with larger chemical shift changes are labeled. They correspond to the putative DNA-binding sites. Helical regions are shown by rectangles at the top of the figure.

Amide ^{1}H and ^{15}N chemical shift differences between the free and bound Mrf-2 are shown in Figure 2. The largest chemical shift changes occur at loop L1, connecting helices H1 and H2. Loop L2 (connecting H4 and H5), helix H5, and the C-terminus also display large chemical shift perturbations. Although secondary effects can cause chemical shift perturbations beyond the direct interacting surface, the sites that form direct protein−DNA contacts generally show the largest changes in chemical shifts. Consequently, the observed chemical shift perturbations suggest that the DNA-binding sites in Mrf-2 are likely to be L1, L2, part of H5, and the C-terminus. This finding is largely consistent with our previous model of DNA recognition, which was based on the structural similarities between the homeodomain and the Mrf-2 ARID in the absence of resonance assignments of the complex (*6*). In this model, H5 contacts the major groove of DNA, and L1 and the C-terminus insert into the flanking minor groove. The large chemical shift changes

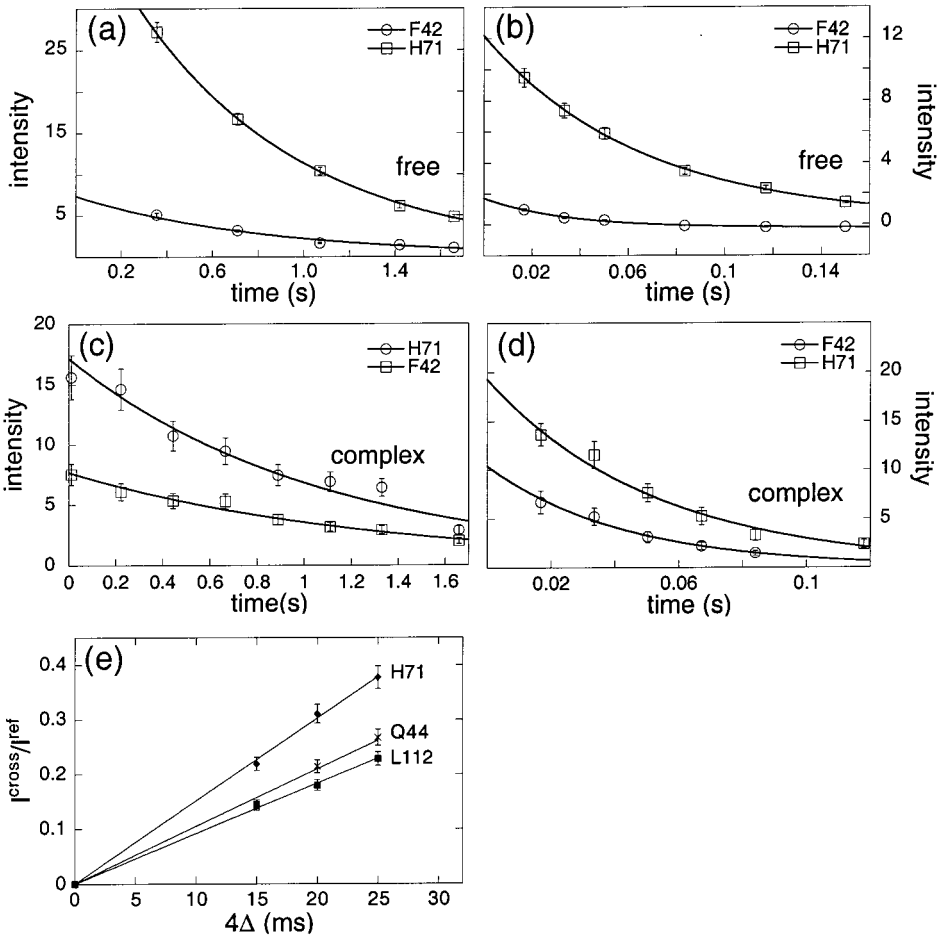

FIGURE 3: Typical curves used for calculations of relaxation rates. Panels (a) and (c) were used for R_1 estimations of the free and bound Mrf-2, respectively; panels (b) and (d) were used for R_2 estimations of the free and the bound Mrf-2, respectively. Panel (e) shows typical curves used for the estimation of CSA/DD cross-correlation relaxation rates of the Mrf-2 domain in complex with DNA. Peak intensities are in arbitrary units.

in L2 suggest that it may also contact the DNA.

^{15}N Relaxation Data. Relaxation rate constants R_1 and R_2 and steady-state NOEs were measured for 85 and 60 amide nitrogen nuclei of the free and the DNA-bound Mrf-2 ARID, respectively (given in the Supporting Information). The CSA/DD cross-correlation rates η_N were measured for 50 residues of Mrf-2 in complex with DNA. All spectra were collected at an ^{15}N frequency of 50.7 MHz. Rate constants of some residues were not measured due to resonance overlap, or intensities that were too weak to be measured accurately. Typical curves for extracting R_1 and R_2 are shown in Figure 3a−d, and those for extracting CSA/DD are shown in Figure 3e. The average R_1 values were 1.42 (with an average error of ±0.07) and 0.92 (with an average error of ±0.04) s^{-1} for the free and DNA-bound Mrf-2, respectively. The average R_2 values were 13.81 (with an average error of ±0.68) and 18.44 (with an average error of ±1.38) s^{-1} for the free and bound Mrf-2 domain, respectively. The changes in the R_1 and R_2 relaxation rates reflect the changes in molecular weight between the free and bound states. The average NOE values were 0.62 (with an average error of ±0.03) for the free Mrf-2 and 0.68 (with an average error of ±0.06) for the DNA-bound Mrf-2. Graphs of relaxation rates versus residue number are shown in Figure 4.

The relaxation data present qualitative information on conformational flexibility. The residues in the helices are more rigid, as indicated by larger NOE values. The residues in L1 and the C-terminus are more flexible than helical regions on a ps−ns time scale in both the free and complexed Mrf-2 ARID, as demonstrated by smaller NOE values. The R_2 values of some residues in L1 of the free Mrf-2 ARID range from approximately 21 to 32 ms, which are abnormally large for a protein of this size, and indicate that these residues may undergo significant motion on a μs−ms time scale. This conformational flexibility is substantially reduced upon DNA binding, consistent with our proposed model that this loop interacts directly with DNA. L2 in the free protein has high mobility, and upon DNA binding, the NOE values demonstrate that this loop becomes more rigid; this observation indicates that L2 is also likely to contact DNA directly. The NOE values in the C-terminal region (residues 115−119) of the free Mrf-2 are negative whereas the NOE values of the same residues in the DNA-bound Mrf-2 are positive; these results suggest that the flexibility of the C-terminus is reduced upon DNA binding, consistent with our model that the C-terminus interacts with DNA.

Diffusion Anisotropy. The model-free approach provides detailed and quantitative information about internal motions; however, diffusion anisotropy of a molecule should be estimated for proper model-free analysis (*18, 19, 20, 23, 24*). The ratio of the long and short axes of the diffusion tensor, $D_\parallel/D_\perp$, of the free Mrf-2 was calculated from the R_2/R_1 ratios of 44 residues (*20, 23, 24*). $D_\parallel/D_\perp$ can also be estimated from longitudinal and transverse dipolar/CSA relaxation interfer-

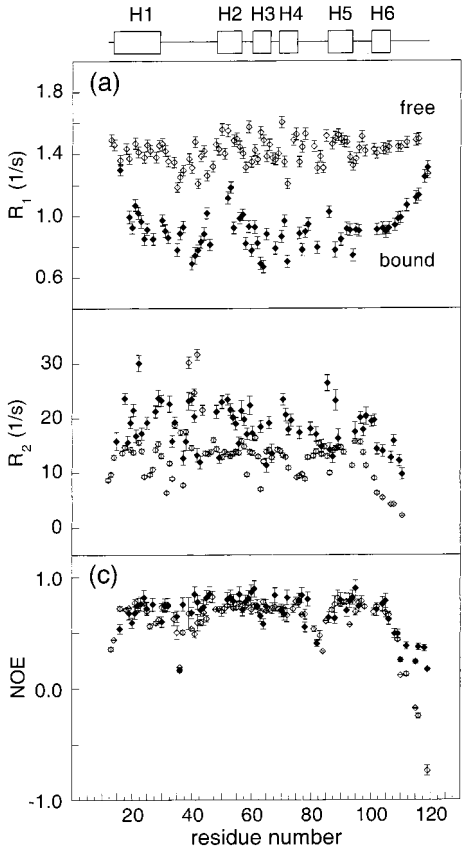

FIGURE 4: ^{15}N relaxation rates (a) R_1 and (b) R_2 and (c) the steady-state ^{15}N-{^{1}H} NOE of Mrf-2 versus residue number. The relaxation rates of the free protein are shown as open diamonds, and those of the complex are shown as solid diamonds.

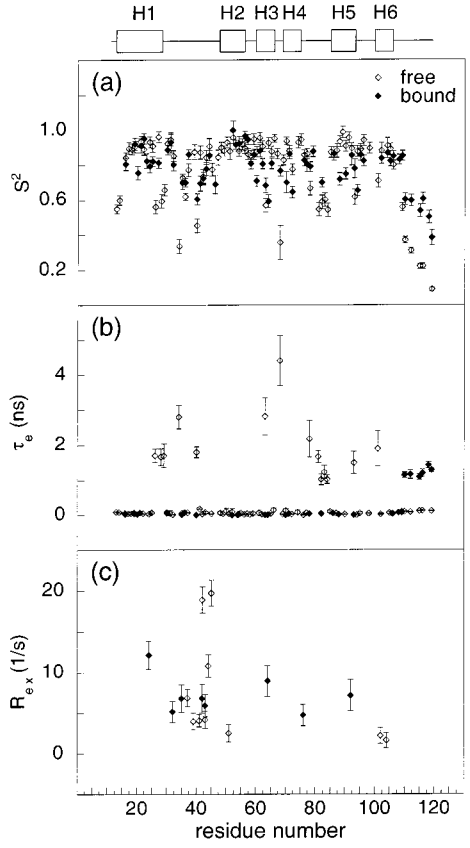

FIGURE 5: Model-free parameters versus residue number of the free and the DNA-bound Mrf-2. The parameters of the free protein are shown as open diamonds, and those for the bound Mrf-2 are shown as solid diamonds. Helical regions are shown by rectangles at the top of the figure. S^2 values generated using individual overall correlation times for the residues in the termini in the free protein (see text) were adjusted according to the overall correlation time by multiplying by τ_c/τ_m, where τ_c is the individual overall correlation time and τ_m is the global correlation time.

ence rates (*25*). The 44 residues are in regions with well-defined structures and also have heteronuclear NOE values of 0.7 or larger. The average $D_{\parallel}/D_{\perp}$ ratio is 1.13 with an average error of $\pm$0.01, using an ensemble of 5 NMR structures. The *Z*-axis of the diffusion tensor could not be defined well in the free Mrf-2 due to the small anisotropy, but is expected to be parallel to the *Z*-axis of the inertia tensor of the molecule. The N−H bond vectors used in the above calculations have a wide angular distribution relative to the *Z*-axis of the inertia tensor of the Mrf-2 molecule (averaged from an ensemble of 5 NMR structures), and the lack of diffusion anisotropy in the free Mrf-2 cannot be due to the lack of data in the calculation. Moreover, the lack of anisotropy is not likely due to low quality of structures, because the set of 5 NMR structures of Mrf-2 has a well-defined backbone conformation with a root-mean-square-deviation from the mean structure of approximately 0.5 Å in the helical regions. The average statistical *F*-test value, representing the goodness-of-fit comparing an isotropic model with an axial symmetric diffusion model, is 0.5; this indicates that an isotropic model is suitable for describing the relaxation data for the free Mrf-2.

For the estimation of the diffusion anisotropy of the Mrf-2−DNA complex, the $D_{\parallel}/D_{\perp}$ ratio was calculated from the R_2/R_1 ratios (*20, 23, 24*) of 27 residues; the residues excluded from this calculation were those with NOE values less than 0.7, in the flexible L1 and L2 loops, in the C-terminal tail, and one residue (Tyr^{24}) suggested to have conformational exchange from CSA/DD cross-correlation rates. The calculation suggests an axial symmetrical property for the complex with the average $D_{\parallel}/D_{\perp}$ ratio from the ensemble of Mrf-2 structures being 1.30 (with an average error of $\pm$0.01). The $D_{\parallel}/D_{\perp}$ ratio of the free or bound Mrf-2 is similar to that estimated from the inertia tensor (*26*) calculated using the Mrf-2 structure or our model of the protein−DNA complex. The *Z*-axis of the diffusion tensor of the Mrf-2−DNA complex is approximately perpendicular to helix H5, and is parallel to the DNA helical axis in our previously proposed model of the protein−DNA complex (*6*).

Model-Free Calculations. To obtain additional quantitative information on the dynamics of the free and the DNA-bound Mrf-2, model-free parameters, S^2, R_{ex}, and τ_e, of the individual residues were extracted from relaxation data (*18, 19*). During model-free calculations, isotropic overall rotation was assumed for both free and bound Mrf-2. Although the diffusion tensor of the Mrf-2−DNA complex suggests a small anisotropy, assuming isotropic rotational diffusion for a system with such small anisotropy should not produce significant errors in the generalized order parameters (*27*). On the other hand, artifacts in R_{ex} and τ_e could occur for bound Mrf-2, and will be addressed later in this section. The overall molecular tumbling times (τ_m) optimized during model-free fitting were 11.6 ($\pm$0.2) ns and 17.4 ($\pm$1.0) ns for the free and bound Mrf-2, respectively.

The motional parameters S^2, τ_e, and R_{ex} of each residue in the free and the complexed Mrf-2 are shown in Figure 5.

The relaxation data were fitted to one of five models in the model-free formalism, where model selection was based on a *F* statistical testing method as described (*21*). The model containing only S^2 was applied to 11 residues in the free Mrf-2 and 28 residues in the bound Mrf-2. A total of 42 residues of the free Mrf-2 and 19 residues of the DNA-bound protein were fitted to the model containing both S^2 and τ_e. The model containing S^2 and R_{ex} was applied to 3 residues in free Mrf-2 and 5 residues in the DNA-bound protein. The model that contains S^2, τ_e, and R_{ex} was used to fit 7 residues of the free Mrf-2 and 2 residues of the bound Mrf-2. Fourteen residues in the free Mrf-2 and 6 residues in the complexed protein require the use of the extended model-free approach for acceptable fits to the relaxation data (*19*). The ^{15}N relaxation rates of all 85 measured residues of the free and 60 of the bound Mrf-2 could be fitted well to the model-free formalism with χ^2 values smaller than the 95% critical value of the χ^2 determined from Monte Carlo simulations.

The last 6 residues at the C-terminus and the first 2 residues at the N-terminus of the free Mrf-2 were fitted to the model-free formalism using the individualized overall correlation time. These residues cannot be fitted with χ^2 values smaller than the 95% critical value with the overall tumbling correlation time of 11.6 ns to any of the five models. These residues are highly flexible as shown by the relaxation data, in particular the residues at the C-terminus, have a random conformation, and are relatively independent from the structured regions. Therefore, their overall correlation times are likely to be different from the structured regions of Mrf-2. These residues were allowed to have individual overall rotational correlation times when fitting the relaxation data to the model containing S^2 and τ_e. The first 2 and the last 6 residues of the free Mrf-2 domain were fitted with local τ_m values ranging from 3.8 to 9.2 ns.

A number of residues in both the free and complexed Mrf-2 require R_{ex} and/or τ_e to have a good fit to their relaxation data in the model-free analysis. Most of these μs−ms and ns internal motions are unlikely to result from artifacts due to neglecting diffusion anisotropy, as rationalized below. For molecules with anisotropic rotational diffusion, the N−H bond vectors that are parallel to the *Z*-axis (unique axis) of the diffusion tensor tend to have the largest R_2/R_1 ratios, and these residues may produce artificial R_{ex} terms in model-free calculations if an isotropic overall tumbling was used. Similarly, the N−H bond vectors that are perpendicular to the *Z*-axis of diffusion tensor tend to have the smallest R_2/R_1 ratios and may result in artificial τ_e terms. Since the free Mrf-2 domain has very small diffusion anisotropy, artifacts are not likely to occur. In addition, the residues in the free Mrf-2 ARID that have R_{ex} terms are mostly located in poorly structured loops where the motions are expected to occur. Most of the residues in bound Mrf-2 that have R_{ex} terms are supported by CSA/DD cross-correlation relaxation rates (*14*, *15*). The reduced spectral density values at zero frequency, $J^{cc}(0)$, can be extracted from the cross-correlation rates, which are free from chemical exchange contributions. Figure 6 shows the plot of $J^{cc}(0)$ and $J(0)$ obtained from R_1, R_2, and NOE relaxation rates for the complexed Mrf-2. The much larger $J(0)$ than $J^{cc}(0)$ indicates exchange contribution. Most of the residues do not show large difference between the two spectral density functions with the average difference between $J^{cc}(0)$ and $J(0)$ of 0.2 ns. Residues Y24, T32, and F42 that have the R_{ex} term from model-free analysis also have larger $J(0)$ than $J^{cc}(0)$. For residues K43 and T64, the values of $J(0)$ and $J^{cc}(0)$ are similar, so that the R_{ex} terms may arise from fitting artifacts. The CSA/DD cross-correlation rates of residues E35 and H92 could not be measured accurately, due to weak intensities.

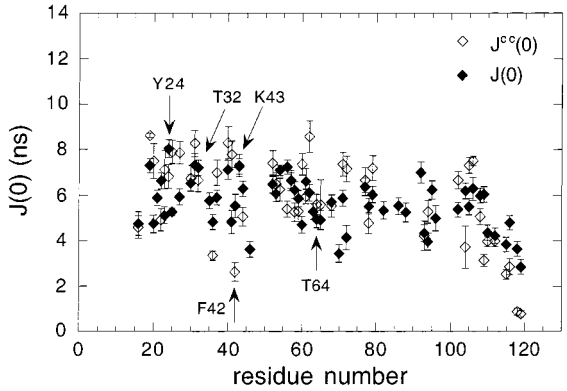

FIGURE 6: Plots of $J^{cc}(0)$ and $J(0)$ for the complexed Mrf-2 at 30 °C. Residues that have R_{ex} terms from model-free analysis and whose cross-correlation rates can be measured are labeled in the figure. The mean values of $J^{cc}(0)$ and $J(0)$ are 5.8 (±0.4) and 5.6 (±0.7) ns, respectively.

Residues with τ_e in the nanosecond range in both free and bound Mrf-2 are mainly in loops (see Figure 5) and have a heteronuclear ^{15}N NOE of 0.6 or smaller, which is indicative of a large amplitude of internal motion. The N−H bond vectors of these residues have a wide angular distribution, and most of them are not perpendicular to the *Z*-axes of the diffusion tensor. Thus, diffusion anisotropy unlikely introduces significant artifacts in τ_e values.

Dynamics of the Two Loops L1 and L2. The two long loops L1 and L2 connecting helices H1 and H2, and helices H4 and H5, respectively, appear to be involved in DNA binding as discussed above. The structures of both loops are not well-defined in the free protein. These two loops, however, undergo different changes in dynamics after complex formation.

L1 clearly experiences dramatic changes in dynamics upon DNA binding. Residues 37, 39, and 41−47 in the free Mrf-2 have large R_{ex} terms ranging from 4.2 (±0.8) to 19.7 (±1.6) s^{-1}, indicating that these residues undergo substantial conformational exchange on the time scale of μs−ms in the free protein. Although this region is flexible on the μs−ms time scale, the dynamics on the ps−ns time scale appear to be quite restricted as shown by large S^2 values. The mean S^2 value of residues 37−47 is 0.79. Residues 34−36 are a flexible segment in this loop with an average S^2 value of 0.56. Upon DNA binding, the dynamics on the μs−ms time scale in L1 become more restricted. Only residues 32, 35, 42, and 43 have R_{ex} terms. The largest R_{ex} value in L1 is reduced from 19.7 (±1.6) s^{-1} for I45 in the free protein to 6.8 (±2.0) s^{-1} for F42 in the complex. Although flexibility in the μs−ms time scale is reduced dramatically in the complex, the dynamics in the ps−ns time scales of this loop do not appear to change upon DNA binding. The average S^2 value of the loop (residues 31−47) is 0.78 (±0.04) in the free Mrf-2 and is 0.76 (±0.04) in the bound Mrf-2.

In contrast to L1, the flexibility of L2 in the ps−ns time scale is reduced upon DNA binding, but the dynamics in the μs−ms time scale are increased. The average S^2 of

residues in L2 is 0.64 (±0.03) for the free protein and increases to 0.79 (±0.04) after complex formation. Residues G78, G81, S82, T83, and S84 required the two-time-scale model for fitting the relaxation data of the free protein. Residues in this region in the protein−DNA complex only require one internal time scale. The average S^2 in L2 in the complex is essentially the same as that of the putative DNA-binding helix H5. The loss of flexibility in the ps−ns time scale in L2 may result from direct contact with the DNA. While the flexibility in L2 decreases on the ps−ns time scale when forming the complex, dynamics on the μs−ms time scale increase upon complex formation. Resonances of several residues (G81, T83, S84, and A85) in this loop are broadened beyond detection.

Dynamic Changes in the C-Terminus. The C-terminus is likely to form important contacts with DNA. Interesting changes in dynamics in this region upon DNA binding are observed. During the model-free analysis, residues in the C-terminus of the free Mrf-2 require the use of individual τ_m that are much smaller than the overall τ_m estimated based on the R_2/R_1 ratios of residues in helical regions. In addition, these residues have small S^2 values. [Mean S^2 is 0.30 (±0.03) for residues from 109 to 119. These S^2 values were adjusted to match the overall global correlation time by multiplying the ratio of the individual correlation time to the τ_m of 11.6 ns.] These results indicate that the C-terminus has a high mobility which is largely independent of the rest of the free protein. This mobility is reduced considerably upon DNA binding. All residues in the C-terminus can fit to the model-free formalism using the same overall correlation time as other residues. However, models with two-time-scale internal motions are necessary for residues 110−119. The S^2 values of the C-terminal residues in the complex have increased with an average value of 0.64 (±0.04) for the last 10 residues. Albeit the conformational flexibility is reduced dramatically, likely due to direct contacts with the DNA, the C-terminus still has significant internal motions, with τ_e in the ns range.

Dynamics of the Helical Regions. Helix H5, a DNA-recognition site as discussed earlier, has a significant reduction in S^2 values upon DNA binding. The average S^2 value of H5 reduced from 0.88 (±0.05) in the free form to 0.77 (±0.04) in the complex. In particular, the decrease of S^2 for residues 88 and 94 is greater than 0.2, which is the average 95% probability difference assuming separable contribution of errors. In addition, several residues in helices H3 and H4 also show a reduction in S^2 values greater than 0.2. Helices 1, 2, and 6, which are not likely to form direct contacts with DNA, have similar S^2 values as those of the free protein as shown in Table 1. The S^2 values of the free and bound Mrf-2 domain are color-coded in the three-dimensional structure shown in Figure 7.

Table 1: Average Order Parameters of Structural Elements in the Free and the DNA-Bound Mrf-2 ARID

structural element	residues included[a]	mean S^2, free form[b]	mean S^2, bound form[c]	difference in S^2 (free − bound)
helix 1	16−29	0.820	0.845	−0.025
helix 1	16−29	0.820	0.845	−0.025
loop 1	31−47	0.781	0.759	0.022
helix 2	48−56	0.906	0.952	−0.046
helix 3	60−66	0.889	0.747	0.142
helix 4	69−75	0.883	0.737	0.146
loop 2	77−84	0.635	0.794	−0.159
helix 5	85−93	0.882	0.775	0.107
kink	95−98	0.922	0.841	0.081
helix 6	101−105	0.849	0.841	0.008
C-terminus[d]	106−119	0.368	0.641	−0.273

[a] Only residues that have relaxation data were used for the calculation of the average S^2 value in each range. [b] Average error in S^2 values of the free Mrf-2 ARID is 5.4%. [c] Average error in S^2 values of the bound Mrf-2 ARID is 6.0%. [d] S^2 values of the C-terminal residues generated using individual overall correlation times in the free protein were adjusted according to the overall correlation time by multiplying by τ_c/τ_m, where τ_c is the individual overall correlation time and τ_m is the global correlation time.

DISCUSSION

Changes of Dynamics in Helical Regions upon Protein−DNA Complex Formation. Although the backbone atoms of helical regions of the Mrf-2 ARID are rigid when free and in complex with DNA, S^2 values decrease for most measured residues in H3, H4, and H5 upon complex formation. It is unlikely that these observations are generated faultily from artifacts in data fitting or differences in experimental conditions. Although it has been shown that aggregation in proteins and protein−ligand complexes may cause artificially larger S^2 values (*27*, *28*), there is no evidence for aggregation of Mrf-2. The free Mrf-2 protein is highly positively charged, as most nucleic acid binding proteins are, and has good solubility. No significant line width changes were observed for samples ranging from 0.2 to 1.5 mM protein concentrations. Mrf-2 in complex with DNA has a solubility less than 1.4 mM (0.6 mM concentration was used in this study), likely due to charge neutralization by complex formation with DNA. Therefore, the potential aggregation of the complex should result in artificially higher but not lower S^2 values, and, thus, the observed decreases in S^2 values in the complex cannot result from plausible aggregation of the complex. The effect of temperature was considered. The relaxation measurements for the free protein were obtained at 26 °C while the complex was measured at 30 °C. The 4 °C changes in temperature should not produce any significant changes in S^2 values, especially for residues at well-structured regions based on previous studies on RNase H (*29*) and a leucine-zipper domain (*30*). Systematic errors in the measurements were also considered. The complex has longer T_1 relaxation times of amide proton and nitrogen nuclei, which can result in errors in NOE measurements. Thus, a longer relaxation delay was used for NOE measurements for the complex than that used for the free protein. In addition, no systematic differences were found between the measured NOEs and that simulated using model-free parameters. Finally, the reductions in average S^2 values in different helices are different and appear to correlate to DNA binding; thus, there does not appear to be a systematic error in S^2 values. Increased mobility throughout a protein upon complex formation has also been observed in other systems such as the mouse major urinary protein bound to its small hydrophobic ligand (*31*) and the C-terminal domain of *E. coli* topoisomerase I bound to a single-strand DNA (*32*). The increase in mobility may partially compensate for the loss of entropy upon binding without significantly perturbing the globular structure of the protein.

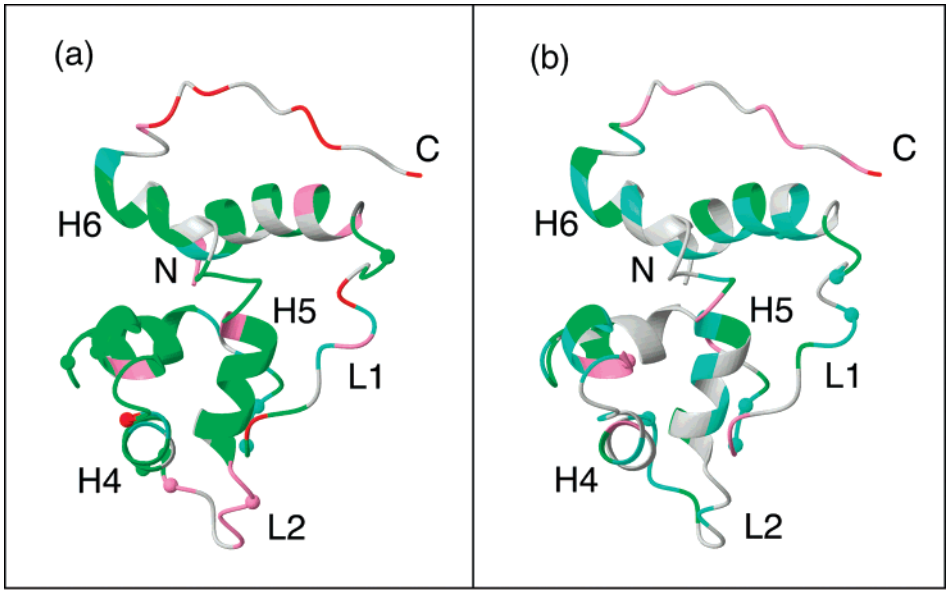

FIGURE 7: Stereoviews of the ribbon diagram of the Mrf-2 ARID showing the changes in dynamics before and after DNA binding. (a) and (b) are colored according to the order parameter S^2 values of the free and the DNA-bound Mrf-2, respectively. The residues with $S^2 \geq 0.85$ (suggesting that they are highly rigid) are shown in green, those with $0.7 \leq S^2 < 0.85$ (relatively rigid) are in cyan, those with $0.7 < S^2 \leq 0.5$ (flexible) are in pink, and those with $S^2 < 0.5$ (highly flexible) are in red. Residues colored in gray are those that could not be measured due to resonance overlap, or the intensities are too weak to be measured accurately. S^2 values generated using individual overall correlation times for the residues in the termini in the free protein (see text) were adjusted according to the overall correlation time by multiplying by τ_c/τ_m, where τ_c is the individual overall correlation time and τ_m is the global correlation time. Balls in the figures represent the residues associated with R_{ex} terms (indicative of motions on the μs−ms time scales). The diagrams were prepared using MOLMOL (*35*).

Conformational Flexibility and Protein−DNA Interaction. Both L1 and L2 have high flexibility, which is a major difference between the Mrf-2 ARID and its homologous DNA-binding domain of Dead ringer (*33*). While neither loop in the Mrf-2 domain is well structured, L1 in Dead ringer forms a β-hairpin. The β-hairpin structure in L1 of Dead ringer is probably stabilized by the interaction with an α-helix formed by the N-terminal residues which are outside of the core homologous region. This helix does not exist in Mrf-2 (*36*).

The flexible loops have reduced mobility upon DNA binding. L1 loses most of the flexibility in the μs−ms time scales upon DNA binding. L2 loses its flexibility on the ps−ns time scale significantly. Dynamic changes are consistent with chemical shift changes suggesting that both loops are involved in direct interactions with DNA. L2 is immediately followed by the putative DNA recognition helix H5. The arrangement of H2, H4, and H5 in Mrf-2 resembles that of helix−turn−helix motifs such as the homeodomain that recognizes DNA (*34*). Therefore, it was logical to propose that H5 of the Mrf-2 ARID, which is the equivalence of the DNA recognition helix of the homeodomain, interacts with the major groove of DNA. This is consistent with the diffusion tensor calculation which shows that H5 orients approximately perpendicular to the unique axis of the diffusion tensor. This long axis is likely parallel to the helical axis of DNA. This was suggested previously for the Mrf-2 ARID (*6*) and also for the homologous DNA-binding domain in Dead ringer (*33*). The adjacent L2 to H5 has large DNA-dependent changes in chemical shifts and flexibility on the μs−ms and ps−ns time scales, suggesting that L2 (which contains 10 residues) is an important element for DNA recognition.

The C-terminus of Mrf-2 is also an important DNA-binding element. The C-terminal segment in the free protein has a random structure and is highly mobile. This mobility is greatly reduced upon binding, likely due to direct contacts with DNA. This direct contact may also cause one of the largest chemical shift perturbations in the C-terminal residues. These results are consistent with the observation that deletion of the C-terminal residues from 109 to 119 abolishes the DNA-binding activity of the Mrf-2 domain, but does not affect the tertiary structure of the protein (unpublished observation).

This is another example that underscores the importance of conformational flexibility in molecular recognition. As in many other cases, the DNA-binding regions of Mrf-2 largely consist of the most flexible segments of the protein. These regions become much less dynamic upon DNA binding, suggesting that they adopt a specific conformation upon complex formation. As implied from Spolar and Record (*1*), changes in flexibility and the induced-fit local conformation should be critical in determining the binding specificity and affinity of the Mrf-2−DNA interaction. Both specificity and affinity are important in transcription regulation and assembly of large nucleoprotein complexes.

In summary, resonance assignments of the backbone and some side-chain atoms for Mrf-2 in complex with DNA have been obtained. Chemical shift data suggest that the Mrf-2 structure in the complex is nearly identical to that of the free protein. Chemical shift perturbations and dynamic studies support the previously proposed model of protein−DNA interactions for Mrf-2; the binding interface with DNA includes L1, L2, H5, and the C-terminus. The flexible loops and the C-terminus have significantly reduced mobility upon interaction with DNA. The decrease in mobility is likely to

result from direct contacts with DNA. Some residues in the putative DNA recognition helix H5, and two other helices, show decreases in S^2 values upon binding to DNA.

ACKNOWLEDGMENT

We thank Dr. Arthur G. Palmer, III, for generously providing us with the programs Modelfree 3.1 and Curvefit. We thank Dr. Geerten W. Vuister for providing us with the pulse sequence for the measurement of ^{15}N CSA/DD cross-correlation rates.

SUPPORTING INFORMATION AVAILABLE

A table is included containing the backbone ^{15}N R_1, R_2 relaxation rates and the heteronuclear NOE values of the Mrf-2 ARID free and in complex with DNA, and the CSA/DD cross-correlation relaxation rates of Mrf-2 ARID in complex with DNA (11 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Spolar, R. S., and Record, M. T., Jr (1994) *Science 263*, 777−784.
2. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−331.
3. Herrscher, R. F., Kaplan, M. H., Lelsz, D. L., Das, C., Scheuermann, R., and Tucker, P. W. (1995) *Genes Dev. 9*, 3067−3082.
4. Gregory, S., Kortschark, R. D., Kalionis, B., and Saint, R. (1996) *Mol. Cell. Biol. 16*, 792−799.
5. Huang, T. H., Oka, T., Asai, T., Okada, T., Merrills, B. W., Gertson, P. N., Whitson, R. H., and Itakura, K. (1996) *Nucleic Acids Res. 24*, 1695−1701.
6. Yuan, Y.-C., Whitson, R. H., Liu, Q., Itakura, K., and Chen, Y. (1998) *Nat. Struct. Biol. 5*, 959−964.
7. Whitson, R. H., Huang, T., and Itakura, K. (1999) *Biochem. Biophys. Res. Commun. 258*, 326−331.
8. Dallas, P. B., Pacchione, S., Wilsker, D., Bowrin, V., Kobayashi, R., and Moran, E. (2000) *Mol. Cell. Biol. 20*, 3137−3146.
9. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−393.
10. Grzesiak, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
11. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson. 101*, 201−205.
12. Marion, D., Ikura, M., and Bax, A. (1989) *J. Magn. Reson. 84*, 425−430.
13. 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.
14. Tessari, M., Mulder, F. A. A., Boelens, R., and Vuister, G. W. (1997) *J. Magn. Reson. 127*, 128−133.
15. Renner, G., and Holak, T. A. (2000) *J. Magn. Reson. 145*, 192−200.
16. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) *J. Biol. NMR 6*, 153−162.
17. Abragam, A. (1961) *Principles of nuclear magnetism*, Clarendon Press, Oxford.
18. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
19. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989−4991.
20. Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995) *J. Am. Chem. Soc. 117*, 12562−12566.
21. Mandel, A. M., Akke, M., and Palmer, A. G., III (1995) *J. Mol. Biol. 246*, 144−163.
22. Wüthrich, K. (1986) *NMR of proteins and Nucleic acids*, Wiley, New York.
23. Brüschweiler, R., Liao, X., and Wright, P. E. (1995) *Science 268*, 886−889.
24. Zheng, Z., Czaplicki, J., and Jardetzky, O. (1995) *Biochemistry 34*, 5212−5223.
25. Kroenke, C. D., Loria, J. P., Lee, L. K., Rance, M., and Palmer, A. G., III (1998) *J. Am. Chem. Soc. 120*, 7905−7915.
26. Copie, V., Tomita, Y., Akiyama, S. K., Aota, S.-I., Yamada, K. M., Venable, R. M., Pastor, R. W., Krueger, S., and Torchia, D. A. (1998) *J. Mol. Biol. 277*, 663−682.
27. Schurr, J. M., Babcock, H. P., and Fujimoto, B. S. (1994) *J. Magn. Reson., Ser. B 105*, 211−224.
28. Liu, Q., Yuan, Y.-C., Shen, B., Chen, D., and Chen, Y. (1999) *Biochemistry 38*, 1415−1425.
29. Mandel, A. M., Akke, M., and Palmer, A. G., III (1996) *Biochemistry 35*, 16009−16023.
30. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., III (1999) *J. Mol. Biol. 305*, 2133−2146.
31. Zidek, L., Novotny, M. V., and Stone, M. J. (1999) *Nat. Struct. Biol. 12*, 1118−1121.
32. Yu, L., Zhu, C. X., Tse-Dinh, Y. C., and Fesik, S. W. (1996) *Biochemistry 35*, 1450−1460.
33. Iwahara, J., and Clubb, R. T. (1999) *EMBO J. 18*, 6084−6094.
34. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) *Cell 67*, 517−530.
35. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph. 14*, 51−55.

BI010476A