Comparison of Simulated and Experimentally Determined Dynamics for a Variant of the LacI DNA-Binding Domain, N*lac*-P

Liskin Swint-Kruse,*# Kathleen Shive Matthews,*# Paul E. Smith,§ and B. Montgomery Pettitt*¶
*The W. M. Keck Center for Computational Biology, #Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005; §Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506; and ¶Institute for Molecular Design, Department of Chemistry, The University of Houston, Houston, Texas 77204-5641 USA

ABSTRACT Recent advances in the experimentally determined structures and dynamics of the domains within LacI provide a rare context for evaluating dynamics calculations. A 1500-ps trajectory was simulated for a variant of the LacI DNA-binding domain, which consists of the first three helices in LacI and the hinge helix of the homologous PurR. Order parameters derived from dynamics simulations are compared to those obtained for the LacI DNA-binding domain with ^{15}N relaxation NMR spectroscopy (Slijper et al., 1997. *Biochemistry.* 36:249–254). The MD simulations suggest that the unstructured loop between helices II and III does not exist in a discrete state under the conditions of no salt and neutral pH, but occupies a continuum of states between the DNA-bound and free structures. Simulations also indicate that the unstructured region between helix III and the hinge helix is very mobile, rendering motions of the hinge helix essentially independent of the rest of the protein. Finally, the α-helical hydrogen bonds in the hinge helix are broken after 1250 ps, perhaps as a prelude to helix unfolding.

INTRODUCTION

Protein structures are dynamic, with motions ranging from large, concerted domain movements to small variations in side chain and backbone positions. The latter motions are probably critical to binding interactions (e.g., Steinbach et al., 1991; Morton and Matthews, 1995) and can contribute to protein stability (e.g., Brooks et al., 1988; Honig and Yang, 1995). As a consequence, successful protein design hinges on our understanding of structural plasticity. However, these high frequency (ps to ns), atomic-level motions can be difficult to monitor experimentally and increasingly are modeled with molecular dynamics (MD) simulations. Reciprocally, the long simulation times needed to overlap most experimentally measured dynamics (μs to ms) are currently infeasible, making verification of the various force field models used for MD calculations difficult. Recent advances in the experimentally determined structures and dynamics of the domains within the *lac* repressor protein (LacI; Lewis et al., 1996; Slijper et al., 1996, 1997; Spronk et al., 1996) provide an opportunity for interpreting and evaluating the accuracy of dynamics calculations.

LacI binding to the *lac* operon is the prototypical system for negative transcriptional control in *Escherichia coli* (for a recent review, see Matthews and Nichols, 1997). The homotetrameric protein has two DNA binding sites, each comprised of the two N-termini of a dimer (Lewis et al., 1996). An isolated, monomeric DNA-binding domain also retains the ability to specifically bind DNA (Geisler and Weber, 1977; Jovin et al., 1977; Ogata and Gilbert, 1979; Khoury et al., 1991). The N-terminus is comprised of three helices (Slijper et al., 1996) which include the classical helix-turn-helix DNA binding motif (Brennan and Matthews, 1989). This domain is attached to the bulkier LacI core through a hinge region (Geisler and Weber, 1977). The small size of the N-terminal fragment (51 amino acids in the DNA binding domain plus 9 residues in the hinge) make it amenable to nuclear magnetic resonance (NMR) studies. Several NMR experiments have determined various dynamic characteristics of this molecule (Slijper et al., 1996, 1997; Spronk et al., 1996).

Most recently, the backbone and side chain dynamics of residues 1–56 have been measured with ^{15}N relaxation. The motions in this region are restricted upon DNA binding, most significantly in the unstructured loop between helices II and III (Slijper et al., 1997). This finding correlates with the previous observation that loop II-III adopts different conformations in DNA-bound and free protein (Slijper et al., 1996). More surprising is the experimental verification that the hinge region, which is a helix when the protein is bound to operator (Lewis et al., 1996; Spronk et al., 1996), is unstructured in the free protein (Spronk et al., 1996; Frank et al., 1997). Similar conformational changes in the hinge region (Schumacher et al., 1994; Nagadoi et al., 1995) have been observed for the highly homologous purine repressor protein (PurR; Weickert and Adhya, 1992).

The kinetics measured via ^{15}N relaxation yield rate constants on the order of 10^8 to 10^{12} per second (e.g., Kay et al., 1989). These motions, and possibly loop II-III movement, are near the time scale of MD simulations (ps to ns). We have performed simulations of the monomeric N-terminus in order to test the current simulation techniques and provide a model of the dynamics responsible for the observed kinetics. The hinge region was included in these studies with hopes of observing events related to helix unfolding.

Received for publication 1 July 1997 and in final form 29 September 1997.
Address reprint requests to Dr. B. Montgomery Pettitt, Dept. of Chemistry, University of Houston, Houston, TX 77204-5641. Tel.: 713-743-3263; Fax: 713-743-2709; E-mail: pettitt@uh.edu.

0006-3495/98/01/413/09 $2.00

Results are compared to those from ^{15}N-NMR relaxation experiments, as well as to the structures of bound and free monomer.

METHODS

The molecular dynamics simulations of the monomeric DNA binding domain of LacI are part of a long-term project to study the effects of oligomerization and DNA binding upon protein structure. The hinge helix was included in the simulations, since several protein-protein and protein-DNA contacts have been postulated to exist in the dimer between the hinge helices of LacI monomers (Lewis et al., 1996; Spronk et al., 1996), and this region apparently undergoes a conformational change upon DNA binding (Nagadoi et al., 1995; Schumacher et al., 1994; Spronk et al., 1996). Since structural coordinates for the side chains of the LacI hinge helices are unavailable, coordinates from the hinge helix of the highly homologous purine repressor protein (Weickert and Adhya, 1992; Schumacher et al., 1994) were substituted to create the chimeric structure used in simulations. The chimeric structure is referred to as "N*lac*-P" and has the sequence shown in Fig. 1. Hinge residues Val-52, Ala-53, and Leu-56 are conserved between LacI and PurR (Weickert and Adhya, 1992). Coordinates for both LacI (Chuprina et al., 1993) and PurR (Schumacher et al., 1994) were taken from structures in which the protein was bound to DNA. By performing the simulations on the chimeric protein in the *absence* of DNA, we allowed for the possibility of simulating any protein structural changes that might occur upon dissociation from the DNA.

The monomer of N*lac*-P used in the present simulation is one-half of a dimer designed for other studies (Swint-Kruse, Matthews, and Pettitt, unpublished results). As a consequence, specific details of the dimer design are included. A Gly-Cys-Gly sequence was appended to a PurR structure (Schumacher et al., 1994) truncated after the histidine residue at the end of the hinge helix, to allow possible dimer formation through disulfide oxidation. The GCG segments were then minimized with ~100 steps of steepest descent while the rest of the construct was restrained. In order to ensure that disulfide bond formation did not disrupt the DNA-bound conformation of the dimer, a grid search of the phi and psi angles of the glycines at position 59 and the C^{α}-C^{β} of the cysteines was performed with Charmm version 22 (Brooks et al., 1983). This exercise demonstrated that the sulfhydryls could be sufficiently close for disulfide bond formation with no distortion of the remaining protein structure. Finally, the GCG tails were rotated to the optimal conformation found by the exhaustive search, the disulfide bond was model-built, and the tails were again minimized with 100 steps of steepest descent.

The next step in the design of N*lac*-P was to align the monomeric DNA binding domain of LacI (DNA-bound form; Chuprina et al., 1993) onto each domain of the PurR-GCG dimer. The point of closest approach of the two proteins was determined by inspection to be the C^{α} of Leu-48 in LacI and the C = O of Ser-46 in PurR. A bond was formed between these two atoms in each DNA-binding domain, and extraneous residues were deleted. The regions of bond formation—LacI 47,48 and PurR 46 C = O,47,48—were minimized with 1500 steps of steepest descent while the rest of the structure was constrained. The chimeric dimer was placed in a box of water (procedure described below) and prepared for dynamics simulations with 20 steps of steepest descent. Monomeric N*lac*-P was extracted from this structure and prepared for MD simulations in its own box of water.

M K P V T_5 L Y D V A_{10} E Y A G V_{15} S Y Q T V_{20} S R V

V N_{25} Q A S H V_{30} S A K T R_{35} E K V E A_{40} A M A E L_{45} N

Y I **P** S_{50} **A V A R** S_{55} **L K V N** H_{60} g c g

FIGURE 1 Sequence of N*lac*-P. The sequence of chimeric protein N*lac*-P is listed using one-letter abbreviations for the amino acids. Sequence numbers are indicated as subscripts. The four helices are overlined, residues from LacI are in regular type, hinge residues from PurR are in bold type, and hinge residues that are conserved between the two parent proteins are underlined. The terminal GCG, designed to allow dimerization through disulfide bond formation, is in lower case.

For efficient calculations with the Ewald sum (De Leeuw et al., 1980; Smith and Pettitt, 1995), which is used to model long-range electrostatic interactions, a balance between box volume and shape must be determined. For example, simulations of a cubic box of water will be faster than those for a rectangular box with equal volume. However, for an elongated protein structure, the volume reduction and subsequent decrease in the number of solvent molecules provided by a rectangular box may outweigh the time advantage of the cubic shape. Monomeric N*lac*-P has dimensions that are far from cubic: 4.2 nm × 3.6 nm × 2.0 nm. Fortunately, the size of the water box needed to accommodate this protein could be reduced *and* better approximate a cube by rotating the protein and placing it along the diagonal of the box. Rotation of 37° around both the *y* and *z* axes was found to be an optimal angle in the case of dimer. This angle was also chosen for monomer, yielding a final box size of 3.7 nm × 3.7 nm × 3.0 nm with an effective protein concentration of 40 mM.

TIP3P water (Jorgensen et al., 1983) was used to solvate the protein structure. Water molecules with oxygen atoms within 2.3 Å of a heavy atom in the protein were deleted. During the following minimizations and trial MD simulations, a few additional water molecules were randomly added to the box so that the pressure of the system was brought to ~1 atm. The final number of waters was 1098; at least three water molecules were between the replicated images of N*lac*-P in all directions. The total number of atoms used in the simulation was 4251.

Molecular dynamics simulations were performed in the presence of a neutralizing background with equal and opposite sign to the total charge on the monomer (Tosi, 1964). We chose this simple model of an ionic atmosphere over explicit ions because the slow diffusion of the latter results in unreasonably long equilibration times (~500 ps; Weerasinghe et al., 1995). The system was subjected to ~50 steps of steepest descent minimization and then "annealed" by first fixing the protein and performing 60 ps of dynamics simulations at 300 K. Next, the solvent was fixed while 5 ps of N*lac*-P dynamics were simulated at 100 K. These two steps were repeated twice more (with only 5 ps simulations in which the protein was fixed) while stepping the temperature of N*lac*-P dynamics 100 K per cycle; a velocity scaling factor "H" was employed at the beginning of each 5 ps, where:

$$H = \sqrt{\frac{T_{\text{new}}}{T_{\text{old}}}} \tag{1}$$

Interactions were described by the all-atom force field of Charmm23 (MacKerell et al., 1992), and the simulation was performed in a microcanonical NVE ensemble. Equations of motion were integrated with a 2 fs time step. The Ewald electrostatic convergence parameter, α, was 1.9 nm^{-1} using all lattice vectors with $n^2 \leq 64$, and the Lennard-Jones and real-space electrostatic cutoff distance was 1.5 nm. Bonds were constrained by the SHAKE algorithm (Ryckaert et al., 1977) with a tolerance of 1×10^{-6} nm. After an additional 60 ps equilibration, a 1500 ps simulation was performed at 300 K. Configurations were saved every 50 steps (0.1 ps).

All simulation data were used in the analyses except for those between 720 and 740 ps, which were lost between data backups. Extensive computer upgrades occurred before data analysis and precluded an exact duplication of this segment of the trajectory. When calculating order parameters, distances moved by various atoms, and root mean square deviations (rmsd) from the simulations, global translational and rotational motions were first removed with a least squares fit of the C^{α} to either the first three helices or to the hinge helix (Smith et al., 1995). The first alignment afforded the best comparison with the NMR data for residues 1–51 of the LacI DNA-binding domain, while the second allowed detailed examination of the hinge helix structure as a function of simulation time.

Order parameters (S^2) were determined from the dynamics data as in Smith et al. (1995), with the following equation (Henry and Szabo, 1985;

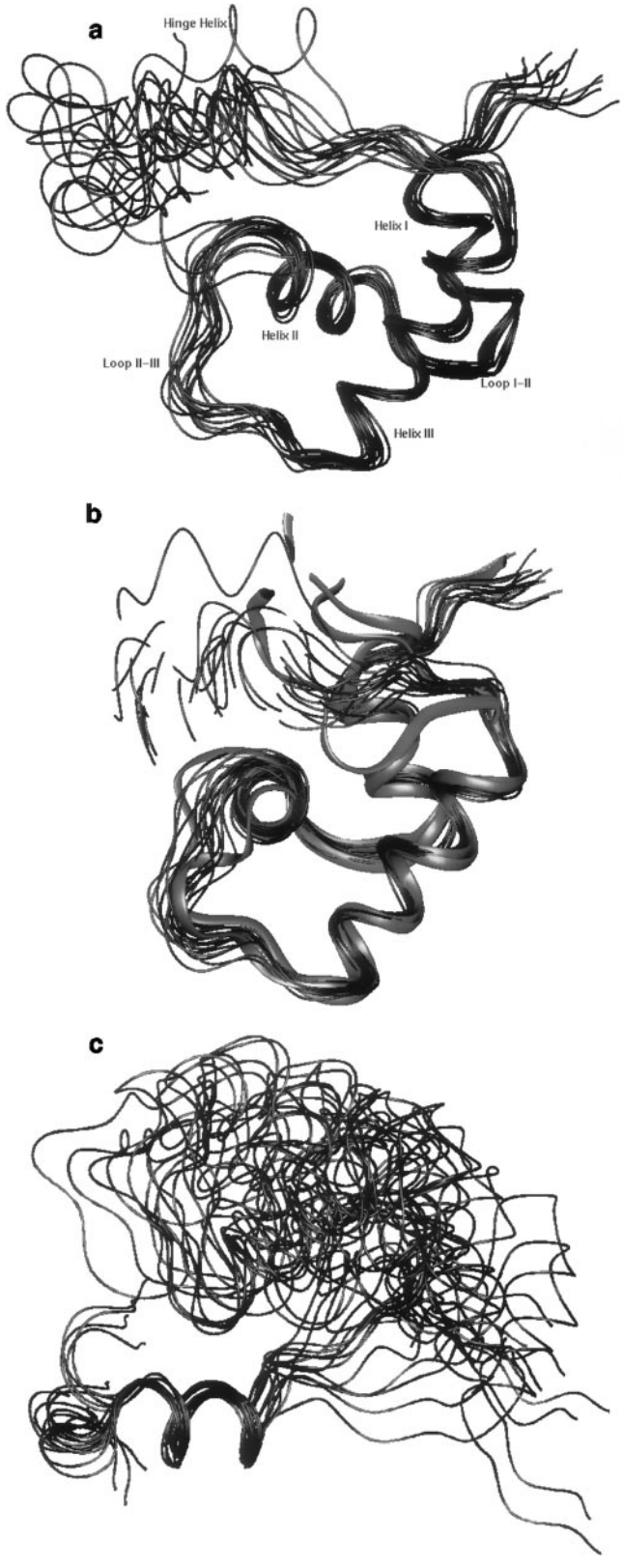

FIGURE 2 Structures generated by the dynamics trajectory. Starting with the structure 40 ps into the trajectory, the C^{α} trace of N*lac*-P is plotted every 100 ps as thin black ribbons. The ribbons were generated with the program Quanta, version 4.1 (Molecular Simulations, Inc., San Diego, CA). (*a*) Alignment along the three helices of the DNA-binding domain. (*b*) The structures in (*a*) are viewed down helix II in order to compare the loop between helices II and III to the experimentally determined structures (*wide, gray ribbons*). The ribbon for the bound structure (Chuprina et al., 1993) is the topmost loop; the ribbon for free LacI is bottommost (Slijper et al., 1996). The ribbons for 340 and 740 ps are not shown in this view. The hinge helix has been cut away in order to better view loop II-III. (*c*) Alignment of structures in view (*a*) along the hinge helix.

Chandrasekhar et al., 1992):

$$S^2 = \frac{1}{2}\left[3\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\langle\mu_\alpha\mu_\beta\rangle^2 - 1\right] \qquad (2)$$

where μ_α ($\alpha = 1, 2, 3$) are the *x*, *y*, and *z* components of the appropriate interatomic unit vector in the macromolecular frame, and the angular brackets denote an ensemble (time) average. Values for internal correlation times were estimated from the first 1420 ps of the correlation function (C_t; Clore et al., 1990; Smith et al., 1995):

$$C_t = S^2 + (1 - S_f^2)\exp(-t/\tau_f) + (S_f^2 - S^2)\exp(-t/\tau_s) \qquad (3)$$

where S_f^2 is a generalized order parameter for a fast motion, τ_f is the correlation time of this motion, and τ_s is a slow correlation time. The generalized order parameter for the latter motion (S_s^2) is implicit in Eq. 3, since $S^2 = S_f^2 * S_s^2$ (Clore et al., 1990). Order parameters and correlation times (τ_e) were estimated from the ^{15}N relaxation data in Fig. 3 of Slijper et al. (1997) using the program Modelfree version 3.1 described by Mandel et al. (1995).

RESULTS AND DISCUSSION

The present dynamics study is part of a long-term project to examine structural changes that occur in the DNA-binding domain and hinge region of LacI upon DNA binding. We have designed a chimeric protein, N*lac*-P, from LacI and PurR that has high-resolution structural coordinates necessary for MD simulations, is small enough for both MD simulations and detailed biophysical studies, and has potential for forming a complete DNA-binding site through disulfide bond formation. The current simulations of monomeric N*lac*-P were accomplished in order to compare the simulated fluctuations to those experimentally determined for the DNA-binding domain of LacI and to ascertain what types of motions are responsible for these kinetics.

Initial analysis of the MD trajectory was achieved by examining overlaid traces of the N*lac*-P C^{α} backbone at 100 ps intervals throughout the trajectory. Results are presented in Fig. 2, *a–c*. Fig. 2 *a* shows that the structures of three helices of the DNA binding domain are well-conserved throughout the simulation. The loop between helices I and II shows only slight changes (Fig. 2 *a*), while loop II-III exhibits much greater mobility. This flexibility is more easily seen in Fig. 2 *b*, where the structures adopted throughout the trajectory by the latter loop (*thin black ribbons*) are compared to the DNA-bound (*top, wide ribbon*; Chuprina et al., 1993) and free structures (*bottom, wide ribbon*; Slijper et al., 1996). The considerable flexibility of this loop predicted by the MD simulations, as compared to

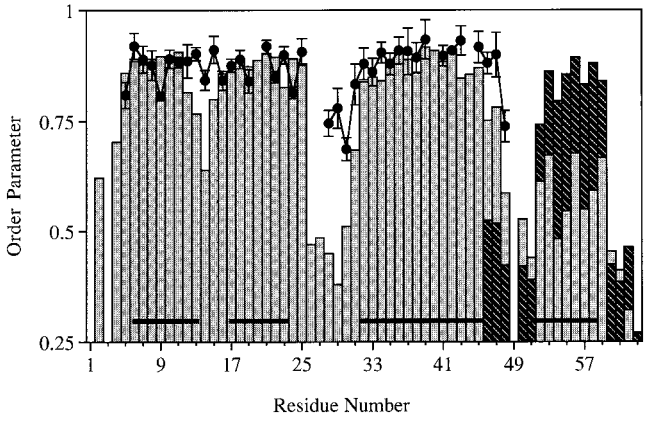

FIGURE 3 Order parameters for backbone NH determined from MD simulations plotted as a function of residue number. Calculations of order parameters are described in the text (Eq. 2). Residues 3 and 49 are prolines and thus S^2 cannot be determined. Black horizontal lines indicate the locations of the four helices. Light gray bars represent values determined after aligning the structures with the first three helices. Values represented with striped bars result when N*lac*-P is aligned along the hinge helix. Circles (●) symbolize values of S^2 estimated from ^{15}N-NMR relaxation data (Slijper et al., 1997; Fig. 3). These data are not included after residue 48, where the sequence of N*lac*-P diverges from the DNA-binding domain of LacI. Error bars represent 95% confidence limits.

the experimentally determined dynamics and structures, is seen in all subsequent data analysis.

A final observation from the structures in Fig. 2 is that the region connecting helix III and the hinge helix is highly mobile, allowing the hinge helix to move essentially independently of the DNA-binding domain. When the structures are aligned on the hinge helix, (Fig. 2 *c*), one observes that the helical structure of this region persists throughout the trajectory, with minor changes occurring only in the last few hundred ps. This observation is further explored later in this section. Although Fig. 2 *c* shows the DNA-binding domain "fanning" out from the hinge region, this motion does not progress from one side of the fan to the other, but moves back and forth within the entire allowed space.

A more detailed picture of structural mobility may be obtained through calculation of generalized order parameters (S^2; Lipari and Szabo, 1982), which reflect the mobility of individual interatomic vectors. These values also allow direct comparison of MD simulations and experimentally determined kinetics. Values of S^2 were calculated for the backbone NH bonds of N*lac*-P from the 1500-ps trajectory. Results are presented in Table 1 and Fig. 3 (*light gray* and *striped* bars). As expected, sites within regular secondary structures (the four helices) have high values of S^2, indicative of low mobility. Several sites within unstructured loops between the helices (residues 14–16, 24–31, and 46–50) have decreased values of S^2, and thus greater mobility. Values for most residues are essentially constant with time (at 500, 1000, and 1500 ps; data not shown), which is expected for a simulation with no gross conformational changes. Residues in loop II-III do show a gradual increase in S^2 (with a total change of $\leq$0.1), reflecting the relaxation of the "DNA-bound" starting structure to the free structure.

Order parameters determined for the DNA-binding domain may be directly compared to S^2 estimated from the ^{15}N-NMR relaxation studies of Slijper et al. (1997; Fig. 3). Simulated and experimental values (Table 1; Fig. 3, *bars* and *circles*, respectively) are in excellent agreement for the three helices of the DNA-binding domain but diverge for the unstructured loops. The MD simulations predict that these loops are more mobile (with lower order parameters) than observed with the ^{15}N relaxation, especially for resi-

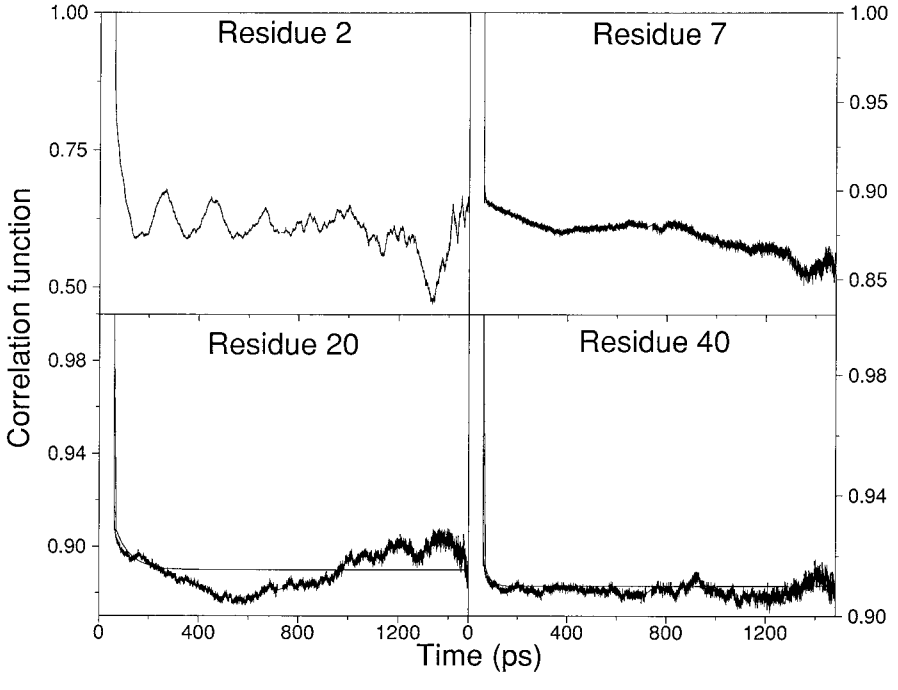

FIGURE 4 Correlation function versus simulation time. Data are shown for four residues (as indicated) which are representative of various types of plots. Thin black lines on the plots for residues 20 and 40 represent the results of fits using Eq. 3.

TABLE 1 General order parameters and relaxation times for N*lac*-P and the DNA-binding domain of LacI

	Molecular Dynamics*					^{15}N-NMR relaxation#	
Residue	S^{2}§	S^{2}¶‖	τ_s (ps)¶	S_f^{2}¶‖	τ_f (ps)¶	S^{2}**	τ_e (ps)**
2	0.62	periodic					
4	0.70	periodic					
5	0.86	0.85	120 (116, 124)##	0.90	0.168 (0.104, 0.259)##	0.81 (0.03)§§	62.6 (29.5)§§
6	0.89	at least three exponential functions required					
7	0.88	at least three exponential functions required				0.89 (0.03)	
8	0.89	0.88	662 (623, 702)	0.90	0.079 (0.067, 0.093)	0.88 (0.03)	91.2 (49.1)
9	0.90	at least three exponential functions required				0.81 (0.01)	43.2 (20.1)
10	0.89	periodic				0.89 (0.02)	71.5 (43.3)
11	0.91	0.91	19.9 (18.2, 21.5)	0.92	0.077 (0.064, 0.091)	0.88 (0.01)	
12	0.82	periodic				0.89 (0.04)	
13	0.77	periodic				0.90 (0.01)	
14	0.64	periodic				0.84 (0.02)	
15	0.80	0.80	4.40 (3.91, 4.91)	0.85¶¶	0.154 (0.132, 0.176)	0.91 (0.03)	
16	0.86	0.86	29.5 (26.5, 32.5)	0.88	0.116 (0.097, 0.137)	0.84 (0.02)	
17	0.86	periodic				0.88 (0.02)	64.6 (64.3)
18	0.89	periodic				0.89 (0.02)	86.1 (45.5)
19	0.87	periodic				0.84 (0.03)	36.5 (9.7)
20	0.89	periodic					
21	0.90	periodic				0.92 (0.01)	130 (79)
22	0.90	0.90	32.6 (30.5, 34.3)	0.91	0.065 (0.048, 0.081)	0.85 (0.01)	92.4 (20.1)
23	0.83	periodic				0.90 (0.02)	
24	0.89	periodic				0.82 (0.01)	48.7‖‖ (16.6)
25	0.88	0.88	38.5 (37.1, 39.7)	0.91	0.105 (0.078, 0.125)	0.91 (0.03)	
26	0.47	complex function					
27	0.49	complex function					
28	0.45	periodic				0.75 (0.03)	194 (58)
29	0.38	periodic				0.78 (0.05)	117 (45)
30	0.51	complex function				0.69 (0.03)	124 (29)
31	0.68	complex function				0.83 (0.05)	500 (624)
32	0.84	periodic				0.88 (0.04)	164 (107)
33	0.85	periodic				0.86 (0.03)	158 (69)
34	0.84	periodic				0.90 (0.02)	
35	0.89	periodic				0.88 (0.02)	71.3 (25.3)
36	0.91	0.91	99.9 (93.4, 105.0)	0.92	0.097 (0.078, 0.113)	0.91 (0.03)	
37	0.89	0.89	86.4 (83.9, 89.2)	0.91	0.088 (0.071, 0.107)	0.91 (0.05)	500 (865)
38	0.86	0.86	43.7 (41.6, 45.7)	0.89	0.087 (0.064, 0.119)	0.89 (0.03)	
39	0.92	0.92	75.6 (72.6, 78.8)	0.93	0.072 (0.055, 0.088)	0.93 (0.04)	
40	0.91	0.91	21.3 (19.4, 23.2)	0.92	0.074 (0.061, 0.086)		
41	0.88	0.88	219 (203, 236)	0.89	0.095 (0.078, 0.114)	0.90 (0.02)	
42	0.91	0.91	319 (306, 331)	0.92	0.085 (0.076, 0.094)	0.91 (0.01)	
43	0.85	complex function				0.93 (0.03)	417 (1693)
44	0.86	0.86	4.34 (2.99, 5.59)	0.87¶¶	0.090 (0.070, 0.112)		
45	0.87	complex function				0.92 (0.04)	
46	0.75	complex function				0.88 (0.02)	
47	0.78	complex function				0.90 (0.05)	225*** (289)
48	0.59	complex function				0.74 (0.04)	89.7### (40.2)
50	0.53	complex function				0.51 (0.01)	113 (15)
51	0.44					0.41 (0.01)	94.1 (10.8)
52	0.61					0.23 (0.02)	500 (77)
53	0.67					0.41 (0.01)	127 (10)

*Determined from the 1500 ps trajectory of N*lac*-P.
#Estimated from Fig. 3 of Slijper et al., 1997.
§Determined using Eq. 2.
¶Determined using Eq. 3.
‖Fitting errors are only in the third or fourth decimal place.
**Determined with Modelfree version 3.1 (Mandel et al., 1995).
##Confidence limits are at one standard deviation.
§§Error at a 95% confidence limit.
¶¶The error of the fit is ±0.01.
‖‖R_{ex} = 0.99 (0.17) s^{-1}.
***R_{ex} = 0.00 (0.51) s^{-1}.
###R_{ex} = 2.61 (0.71) s^{-1}.

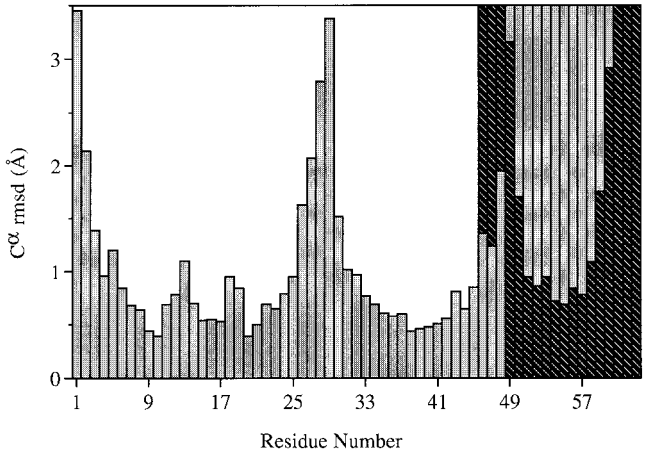

FIGURE 5 Root mean square deviations of C^{α} during MD simulations plotted as a function of residue number. Light gray bars represent values determined after aligning the structures with the first three helices. Values represented with striped bars result when N*lac*-P is aligned along the hinge helix.

dues in loop II-III. This difference may be due to either artifacts of the simulation or limitations in analysis of the NMR data. However, several lines of reasoning suggest that the discrepancy reflects a real deviation, probably arising from the significant differences in the solution conditions of the two techniques. First, the short times of dynamics simulations usually *under*-predict mobility, the opposite of the present results. Second, the range of motion observed in the simulation stays near the boundaries set by the two experimental structures. Finally, changes in salt concentration are known to have significant effects on the stability of this protein: NMR experiments were performed at 400 mM KCl with 60 mM potassium phosphate at pH 4.5 (Slijper et al., 1996, 1997), while resource-constraints dictated that MD calculations be carried out at zero ionic strength and neutral pH. The midpoint of thermal denaturation for the DNA-binding domain of LacI decreases nearly 20°C as the salt concentration is lowered from 500 to 0 mM, with a concomitant increase in susceptibility to tryptic digestion (Schnarr and Maurizot, 1982). Decreased protein stability for staphylococcal nuclease, thioredoxin, and their mutants has been correlated with an increase in the number of residues with low order parameters (Alexandrescu et al., 1996; De Lorimier et al., 1996).

An opposite trend with changing stability has been observed for correlation times (τ), which report the time constants of the motions with amplitudes described by S^2 (De Lorimier et al., 1996). However, a direct comparison of correlation times from the current simulation and ^{15}N-NMR experiments could not be accomplished. Results from the simulation are presented in Table 1 and Fig. 4. Most of the data are too complex to be fit with Eq. 3. Fig. 4 shows representative plots for several of these behaviors: residues 2 and 20 appear to have a periodic recurrence superimposed on the double exponential decay, while residue 7 requires at least a triple exponential decay function. Other behaviors are even more varied. The fit of data for residue 40 is typical of those sites which can be reasonably described by Eq. 3. However, correlation times for these residues still cannot be compared directly to experimentally determined values, since the number of parameters that may be estimated from the NMR data are severely restricted by having only three data points (Alexandrescu et al., 1996). Even for the four-

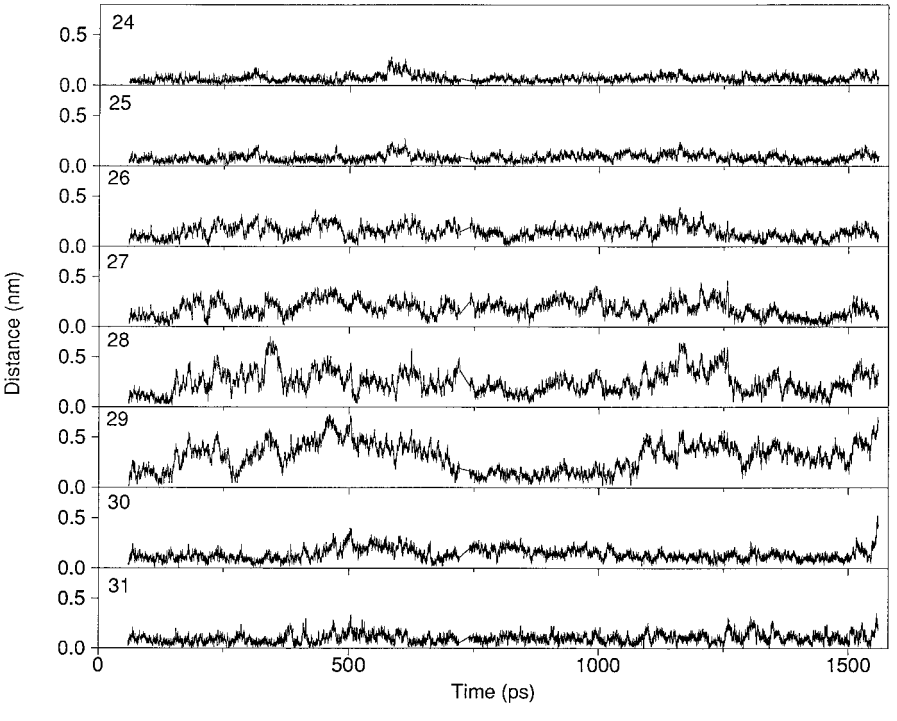

FIGURE 6 Distance of C^{α} in the loop between helices II and III from the DNA bound structure versus simulation time. Residue numbers for the C^{α} in loop II-III are in the top left-hand corners of their respective plots. Values close to zero indicate that the C^{α} is close to the position of the DNA bound form of the protein (Chuprina et al., 1993).

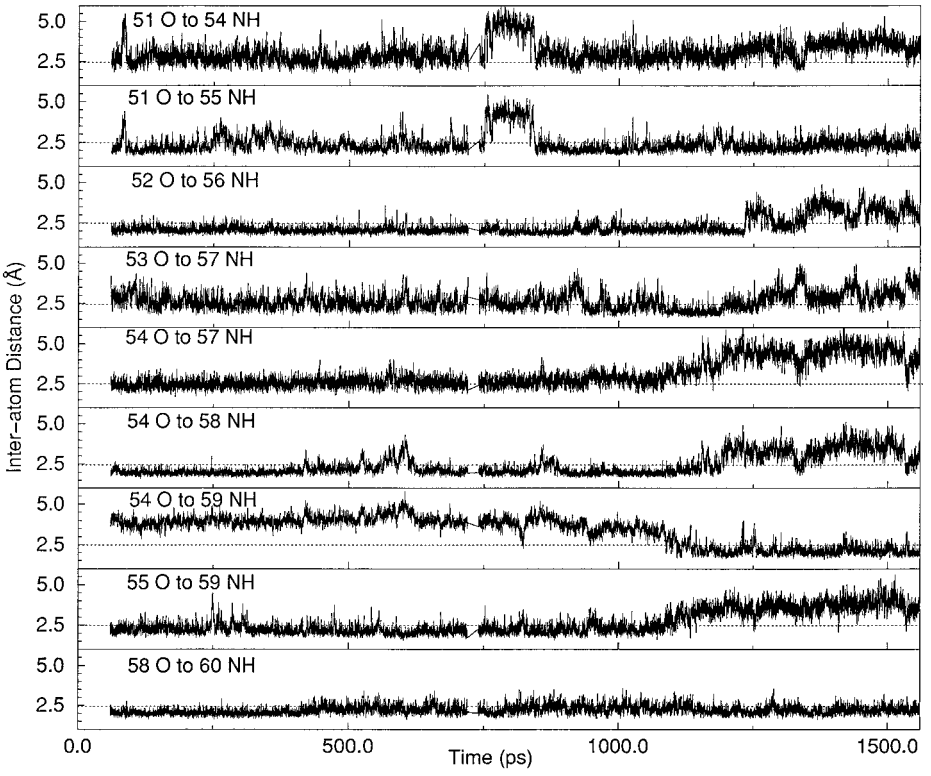

FIGURE 7 Interatomic distances for atoms involved in hydrogen bonds of the hinge helix as a function of simulation time. Donor and acceptor atoms of the hydrogen bonds are listed in the top left-hand corners of their respective plots. The dotted lines at 2.5 Å represent an average hydrogen bond length and are included to aid inspection of the data.

residue subset of sites for which both simulated and experimental correlation times can be determined (Table 1), the NMR data are limited to a single value of τ_e, while the simulated data require both τ_f and τ_s for a complete description.

The motional independence of the DNA-binding domain and the hinge helix seen in Fig. 2 *c* is again evident when values of S^2 for the hinge are calculated from the MD simulation with different initial alignments: along the DNA-binding domain (Fig. 3, *light gray bars*) and with the hinge helix (Fig. 3, *striped bars*). Values increase in the latter alignment, showing that the structure is conserved though not quite as rigidly as helices I–III. This behavior is echoed in Fig. 5, where the C^α rmsd is plotted for each residue using the two possible alignments for the hinge helix calculations. The rigid body motion of the DNA-binding domain relative to the hinge may contribute to the factors that render this region of LacI undetectable with x-ray diffraction techniques in the absence of DNA (Lewis et al., 1996).

The overall pattern of the C^α rmsd in Fig. 5 is strikingly similar to that generated by the 32 structures of the free DNA-binding domain reported by Slijper et al. (1996; Fig. 2 *b*). Although the magnitudes reported on the two plots are different for loop II-III (see above), both data sets show that residues 28 and 29 have the highest mobilities apart from the N and C termini. The preliminary analysis presented in Fig. 2 *b* shows a broad continuum of structures for this loop rather than two discrete states as reported by Slijper et al. (1996) based on the NMR data. Therefore, a more detailed examination was performed by monitoring the distance of loop II-III C^α from the starting, DNA-bound structure (Chuprina et al., 1993) during the course of the simulation trajectory (Fig. 6). Under the solution conditions of the simulation, none of the residues in loop II-III occupy a discrete, DNA-free state. Several of the residues, most notably His-29, do appear to spend time in a state very similar to the DNA-bound structure; an example of this behavior is seen for His-29 between 750 and 1000 ps.

We concluded our analysis of the MD simulation by looking for changes in the hinge helix (with PurR sequence) upon DNA dissociation. Experimental evidence (Nagadoi et al., 1995; Spronk et al., 1996; Frank et al., 1997) supports the hypothesis that this helix unwinds in the absence of DNA for LacI, PurR, and other homologous proteins (Ha et al., 1989; Frank et al., 1997). Visual inspection of structures throughout the trajectory shows only subtle changes in this region of N*lac*-P (Fig. 2 *c*). This finding is not surprising, given that Williams et al. (1996) have measured a time constant for helix unfolding of 160 ns, two orders of magnitude longer than the simulation time of this trajectory. However, plots of interatomic distance versus simulation time for atoms involved in hydrogen bonds are more revealing (Fig. 7). Around 1250 ps of simulation, the i to i+3 hydrogen bonds between 51 and 54 and between 54 and 57 are lost, as are the i to i+4 bonds between 52 and 56, 53 and 57, 54 and 58, and 55 and 59 (Fig. 7). Conversely, none of the α-helical bonds are lost in helices I–III (data not shown). Hydrogen bonds toward the C terminus of the hinge helix are lost earlier than those near the N terminus. At this same time, a non-α-helical, i to i+5 bond is gained between

residues 54 and 59 (Fig. 7). These new patterns are maintained throughout the remainder of the simulation trajectory and may be a prelude to unwinding after DNA dissociation.

In conclusion, the comparison between simulated and experimentally determined dynamics shows good agreement, especially for the generalized order parameters in well-ordered regions of protein structure. The discrepancies observed in the unstructured loops could be due to decreased protein stability resulting from differences in the solution conditions of the two techniques. Taken together, these results indicate that the time needed for an even more rigorous test of the simulation algorithm using a full-scale model (with ions) would be well-spent. Further experimental examination of the relationship between dynamics and global protein stability also promises to be very illuminating.

The authors thank Dr. Gillian Lynch (The University of Houston) for assistance with the FORTRAN programming necessary for data analysis and Dr. Edward P. Nikonowicz and Eric DeJong (Rice University) for many helpful discussions on calculating order parameters from ^{15}N-NMR relaxation data and access to their copy of Modelfree (version 3.1; Mandel et al., 1995).

L.S.K. is supported by National Library of Medicine Postdoctoral Training Grant 1-T15-LM-07093 in the W. M. Keck Center for Computational Biology. This work was also supported by National Institutes of Health Grant GM22441 and Robert A. Welch Foundation Grant C-576 (to K.S.M.); grants from the National Institutes of Health, National Science Foundation, and Robert A. Welch Foundation (to B.M.P.); and by the Kansas Agricultural Experimental Station (P.E.S.; Publication 98-67-J).

REFERENCES

Alexandrescu, A. T., W. Jahnke, R. Wiltscheck, and M. J. Blommers. 1996. Accretion of structure in staphylococcal nuclease: an ^{15}N-NMR relaxation study. *J. Mol. Biol.* 260:570–587.

Brennan, R. G., and B. W. Matthews. 1989. The helix-turn-helix DNA binding motif. *J. Biol. Chem.* 264:1903–1906.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.

Brooks, C. L. III, M. Karplus, and B. M. Pettitt. 1988. Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. Adv. Chem. Phys., Vol 71. John Wiley & Sons, New York.

Chandrasekhar, I., G. M. Clore, A. Szabo, A. M. Gronenborn, and B. R. Brooks. 1992. A 500 ps molecular dynamics simulation study of interleukin-1β in water. *J. Mol. Biol.* 226:239–250.

Chuprina, V. P., J. A. C. Rullmann, R. M. J. N. Lamerichs, J. H. van Boom, R. Boelens, and R. Kaptein. 1993. Structure of the complex of *lac* repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. Mol. Biol.* 234:446–462.

Clore, G. M., A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn. 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112:4989–4991.

de Leeuw, S. W., J. W. Perram, and E. R. Smith. 1980. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. Lond. A.* 373:27–56.

De Lorimier, R., H. W. Hellinga, and L. D. Spicer. 1996. NMR studies of structure, hydrogen exchange, and main-chain dynamics in a disrupted-core mutant of thioredoxin. *Protein Sci.* 5:2552–2565.

Frank, D. E., R. M. Saecker, J. P. Bond, M. W. Capp, O. V. Tsodikov, S. E. Melcher, M. M. Levandoski, and M. T. Record. 1997. Thermodynamics of the interactions of *lac* repressor with variants of the symmetric *lac* operator: effects of converting a consensus site to a nonspecific site. *J. Mol. Biol.* 267:1186–1206.

Geisler, N., and K. Weber. 1977. Isolation of the amino-terminal fragment of lactose repressor necessary for DNA binding. *Biochemistry*. 16: 938–943.

Ha, J.-H., R. S. Spolar, and M. T. Record, Jr. 1989. Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J. Mol. Biol.* 209:801–816.

Henry, E. R., and A. Szabo. 1985. Influence of vibrational motion on solid state line shapes and NMR relaxation. *J. Chem. Phys.* 82:4753–4761.

Honig, B., and A. S. Yang. 1995. Free energy balance in protein folding. *Adv. Protein Chem.* 46:27–58.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Jovin, T. M., N. Geisler, and K. Weber. 1977. Amino-terminal fragments of *Escherichia coli lac* repressor bind to DNA. *Nature*. 269:668–672.

Kay, L. E., D. A. Torchia, and A. Bax. 1989. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*. 28: 8972–8979.

Khoury, A. M., H. S. Nick, and P. Lu. 1991. In vivo interaction of *Escherichia coli lac* repressor N-terminal fragments with the *lac* operator. *J. Mol. Biol* 219:623–634.

Lewis, M., G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*. 271:1247–1254.

Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.

MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, D. T. Nguyen, T. Ngo, B. Prodhom, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, J. Wiorkiewicz-Kuczera, and M. Karplus. 1992. Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *FASEB J.* 6:A143.

Mandel, A. M., M. Akke, and A. G. Palmer III. 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246:144–163.

Matthews, K. S., and J. C. Nichols. 1997. Lactose repressor protein: functional properties and structure. *In* Progress in Nucleic Acids Research and Molecular Biology. K. Moldave, editor. Academic Press, San Diego. In press.

Morton, A., and B. W. Matthews. 1995. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry*. 34:8576–8588.

Nagadoi, A., S. Morikawa, H. Nakamura, M. Enari, K. Kobayashi, H. Yamamoto, G. Sampei, K. Mizobuchi, M. A. Schumacher, R. G. Brennan, and Y. Nishimura. 1995. Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. *Structure*. 3:1217–1224.

Ogata, R. T., and W. Gilbert. 1979. DNA-binding site of *lac* repressor probed by dimethylsulfate methylation of *lac* operator. *J. Mol. Biol.* 132:709–728.

Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327–341.

Schnarr, M., and J.-C. Maurizot. 1982. Stability of the *lac* repressor headpiece against thermal denaturation and tryptic hydrolysis. *Biochim. Biophys. Acta*. 702:155–162.

Schumacher, M. A., K. Y. Choi, H. Zalkin, and R. G. Brennan. 1994. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science*. 266:763–769.

Slijper, M., R. Boelens, A. L. Davis, R. N. H. Konings, G. A. van der Marel, J. H. van Boom, and R. Kaptein. 1997. Backbone and side chain dynamics of *lac* repressor headpiece (1–56) and its complex with DNA. *Biochemistry*. 36:249–254.

Slijper, M., A. M. J. J. Bonvin, R. Boelens, and R. Kaptein. 1996. Refined structure of *lac* repressor headpiece (1–56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the *lac* operator. *J. Mol. Biol.* 259:761–773.

Smith, P. E., and B. M. Pettitt. 1995. Efficient Ewald electrostatic calculations for large systems. *Comput. Phys. Comm.* 91:339–344.

Smith, P. E., R. C. van Schaik, T. Szyperski, K. Wüthrich, and W. F. van Gunsteren. 1995. Internal mobility of the basic pancreatic trypsin inhibitor in solution: a comparison of NMR spin relaxation measurements and molecular dynamics simulations. *J. Mol. Biol.* 246:356–365.

Spronk, C. A. E. M., M. Slijper, J. H. van Boom, R. Kaptein, and R. Boelens. 1996. Formation of the hinge helix in the *lac* repressor is induced upon binding to the *lac* operator. *Nat. Struct. Biol.* 3:916–919.

Steinbach, P. J., A. Ansari, J. Berendzen, D. Braunstein, K. Chu, B. R. Cowen, D. Ehrenstein, H. Frauenfelder, J. B. Johnson, D. C. Lamb, S. Luck, J. R. Mourant, G. U. Neinhaus, P. Ormos, R. Phillipp, A. Xie, and R. D. Young. 1991. Ligand binding to heme proteins: connection between dynamics and function. *Biochemistry*. 30:3988–4001.

Tosi, M. P. 1964. Cohesion of ionic solids in the Born model. *In* Solid State Physics, Vol. 16. F. Seitz and D. Turnbull, editors. Academic Press, New York. 1–120.

Weerasinghe, S., P. E. Smith, V. Mohan, Y.-K. Cheng, and B. Montgomery Pettitt. 1995. Nanosecond dynamics and structure of a model DNA triple helix in saltwater solution. *J. Am. Chem. Soc.* 117:2147–2158.

Weickert, M. J., and S. Adhya. 1992. A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* 267: 15869–15874.

Williams, S., T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer. 1996. Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry*. 35: 691–697.