11. R. G. Burns, *Mineralogical Application of Crystal Field Theory* (Cambridge Univ. Press, Cambridge, 1993).
12. E. S. Gaffney, D. L. Anderson, *J. Geophys. Res.* **78**, 7005 (1973).
13. D. M. Sherman, *J. Geophys. Res.* **96**, 14299 (1991).
14. F. Sette *et al.*, *Phys. Rev. Lett.* **75**, 850 (1995).
15. See supporting data on *Science* Online.
16. J.-P. Rueff *et al.*, *Phys. Rev. B* **60**, 14510 (1999).
17. J. Badro *et al.*, *Phys. Rev. Lett.* **83**, 4101 (1999).
18. J.-P. Rueff *et al.*, *Phys. Rev. Lett.* **82**, 3284 (1999).
19. G. Peng *et al.*, *Appl. Phys. Lett.* **65**, 2527 (1994).
20. J. Li *et al.*, *Eos* **84** (suppl.), 46 (abstr. S21E-0376) (2003).
21. G. Vankó *et al.*, *ESRF Highlights* **2002**, 59 (2002); available online at http://www.esrf.fr/UsersAndScience/Publications/Highlights/2002/HRRS/HRRS9/
22. C. A. McCammon, *Nature* **387**, 694 (1997).
23. J. Zhang, D. J. Weidner, *Science* **284**, 782 (1999).
24. I. Daniel *et al.*, *Geophys. Res. Lett.* **28**, 3789 (2001).
25. D. Andrault, N. Bolfan-Casanova, N. Guignot, *Earth Planet. Sci. Lett.* **193**, 501 (2001).
26. P. Gillet *et al.*, *Phys. Earth Planet. Inter.* **117**, 361 (2000).
27. J. M. Brown, T. J. Shankland, *Geophys. J. R. Astron. Soc.* **66**, 579 (1981).
28. M. Murakami *et al.*, *Science* **304**, 855 (2004).
29. H. Keppler, C. A. McCammon, D. C. Rubie, *Am. Min.* **79**, 1215 (1994).
30. Q. Williams *et al.*, *J. Geophys. Res.* **95**, 21549 (1990).
31. R. Boehler, *Rev. Geophys.* **38**, 221 (2000).
32. The CCD camera is slightly sensitive in the IR, which allows imaging of the laser spot itself.
33. R. D. van der Hilst, H. Kárason, *Science* **283**, 1885 (1999).
34. A. M. Hofmeister, *Science* **283**, 1699 (1999).

Supporting Online Material
www.sciencemag.org/cgi/content/full/305/5682/383/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
References and Notes

6 April 2004; accepted 10 June 2004

Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes

Charalampos G. Kalodimos,[1]* Nikolaos Biris,[1]* Alexandre M. J. J. Bonvin,[1] Marc M. Levandoski,[2] Marc Guennuegues,[1] Rolf Boelens,[1] Robert Kaptein[1]†

Interaction of regulatory DNA binding proteins with their target sites is usually preceded by binding to nonspecific DNA. This speeds up the search for the target site by several orders of magnitude. We report the solution structure and dynamics of the complex of a dimeric *lac* repressor DNA binding domain with nonspecific DNA. The same set of residues can switch roles from a purely electrostatic interaction with the DNA backbone in the nonspecific complex to a highly specific binding mode with the base pairs of the cognate operator sequence. The protein-DNA interface of the nonspecific complex is flexible on biologically relevant time scales that may assist in the rapid and efficient finding of the target site.

Protein-nucleic acid interactions are responsible for the regulation of key biological functions such as transcription, translation, replication, and recombination. Understanding the mechanisms by which regulatory proteins discern their target sequences within the DNA genome requires that we also understand the properties of their complexes with nonspecific DNA (*1*). Nonspecific sites participate in the regulation of the physiological function because they complex, in vivo, most of the DNA binding protein molecules that are not bound at their regulatory functional sites (*2*). Furthermore, protein–nonspecific DNA interactions may also play an important role in the in vivo translocation of DNA binding proteins (*3*, *4*). Indeed, it has been demonstrated that proteins can find their DNA target sites at rates that are much faster than diffusion-controlled by initially binding to DNA anywhere along the chain and then translocating to their specific sites by a combination of intramolecular processes, including one-dimensional (1D) diffusion along the DNA (*5*–*7*). Most, if not all, proteins that interact with specific sites bind also nonspecifically to DNA with appreciable affinity. Thus, nonspecific interaction is an important intermediate step in the process of sequence-specific recognition and binding (*8*). Description of the structural and dynamic response of protein binding to both nonspecific and cognate operator sequences is a prerequisite for elucidating the physicochemical mechanisms that underlie the protein-DNA recognition process.

Many studies have revealed that sequence-specific binding is coupled to extensive conformational changes in both protein and DNA components (*9*, *10*). How does protein binding to nonspecific DNA, which precedes the specific binding, alter the structural and dynamic features of the partners? How does the complex switch from the nonspecific to the specific mode during the exploration process? These questions remain unanswered, mainly because of the scarcity of data on nonspecific binding. To address these central issues, we have structurally and dynamically characterized the DNA binding domain (DBD) of the lactose repressor in the free state and bound to a nonspecific DNA sequence and to its natural operator. *Lac* repressor, which is a prototype of transcription regulation, binds with a dimeric DNA binding unit to its cognate operator sequences 10^2 to 10^3 times faster than the rate estimated for a 3D diffusion-controlled reaction, with very high specificity (*11*). We previously reported the structure of the dimeric *lac* DBD free in solution and bound to its natural operator *O1* (*12*, *13*). Here, we report the high-resolution structure of the dimeric *lac* DBD bound to an 18-base-pair nonspecific DNA fragment (NOD). The variation of flexibility and the redistribution of the native-state ensemble along the recognition pathway were probed by relaxation and hydrogen-deuterium (H-D) exchange methodologies. The new findings also provide a molecular basis for explaining the thermodynamics of specific compared with nonspecific binding.

A dimeric *lac* headpiece-62 (HP62) mutant (*12*, *14*), which binds DNA with an affinity similar to that of the intact *lac* repressor, formed a stable and unique complex with an 18-base-pair-long nonspecific fragment (fig. S1) (*15*). The structure was determined on the basis of 2412 experimental restraints derived from multidimensional nuclear magnetic resonance (NMR) spectroscopy [Supporting Online Material (SOM) Text and table S1]. The ensemble of the 20 lowest-energy conformers is depicted in Fig. 1A.

Remarkably, the protein tilts by ~25°, relative to the DNA (Fig. 2A), compared to the specific complex, resulting in a dramatic alteration of the protein-DNA contacts (Fig. 2C). The side chains of Tyr^7, Tyr^{17}, Gln^{18}, and Arg^{22}, the primary residues that confer specificity to *lac* repressor through direct interactions with the base pairs in the major groove of the cognate operator (Fig. 2C) (*13*), shift and twist so as to participate in hydrogen bonds or electrostatic interactions with the phosphates (Figs. 1B and 2B). Consequently, a cavity is formed in the protein-DNA interface that can accommodate water molecules. In some of the structural conformers, these residues interact with the major groove through water bridges (SOM Text and fig. S3). A direct intermolecular hydrogen bond between the side chain of Tyr^{17} and the O2P phosphate atom of Thy^8 is present in all conformers. We tested the observed interaction of the Tyr^{17} hydroxyl group with the phosphate backbone by examining the DNA binding of a mutant repressor in which Tyr^{17} was mutated to Phe (Y17F) in the context of full-length *lac* repressor (SOM Text). For binding to the specific operator, removal of the Tyr^{17} OH

[1]Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, Netherlands. [2]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA.

*Present address: Department of Chemistry, Rutgers University, Newark, NJ 07102, USA.
†To whom correspondence should be addressed. E-mail: kaptein@nmr.chem.uu.nl

group results in ~100-fold reduction in affinity (Fig. 2D). Notably, a ~10-fold reduction in affinity relative to wild-type repressor was observed for binding of Y17F repressor to nonspecific DNA. These results are in excellent agreement with our structural data, which highlight the dual role of the recognition helix residues: They confer specificity in the cognate complex, whereas in a different role they increase the stability of the nonspecific complex.

The rearrangement of the protein-DNA surface causes all the base pair interactions that are present in the specific complex to be lost in the nonspecific complex throughout the recognition sequence (Fig. 2C). The nonspecific mode is almost totally electrostatic, as was earlier suggested on the basis of thermodynamic analysis (*16*). For example, the side chain of Arg^{22} is reoriented and in ~60% of the conformers it makes electrostatic and/or hydrogen bond contacts with the phosphate backbone. Interestingly, Arg^{22} is also seen to interact with the NH_2 groups of the adenines Ade3 and Ade15 through water molecules (fig. S3). In contrast, in the specific complex Arg^{22} interacts uniquely and directly with the bases in the major groove and lies more than 6 Å apart from the phosphates. Compared to the specific complex, two additional charged residues, Arg^{35} and Lys^{37}, approach the DNA phosphates in the nonspecific complex within a Coulombic interaction distance (<6 Å), thus enhancing its electrostatic character (Fig. 2C and SOM Text). Numerous protein side-chain and backbone contacts to the sugar phosphates further stabilize the nonspecific complex (Fig. 1B). Interestingly, most of these interactions are preserved in the specific complex (Fig. 2C). The persistence of these intermolecular contacts may be energetically important for a smooth switch from the nonspecific to the specific binding mode when the cognate sequence is encountered during the exploration process. Thus, the protein presents a scaffold for DNA binding that can be fine-tuned for either nonspecific or specific binding.

Inducer binding relieves *lac* repression by reducing its DNA binding affinity to nonspecific levels. The hinge region (residues 50 to 58) is postulated to hold a central role in the induction mechanism (*17*–*19*). This region is disordered in the free state, but it forms an α helix when bound to operator DNA (Fig. 3). Binding of *lac* DBD to nonspecific sequences does not induce α helix formation in the hinge region (Figs. 1 and 3), further emphasizing the importance of this region both for specificity and as a structural switch between nonspecific and specific binding modes. In its unfolded state, the hinge region makes no contacts with the DNA minor groove in the nonspecific complex, and thus DNA remains in the canonical B form instead of having a central kink as observed in the specific complex (*20*) (Figs. 1 and 3). The observation that neither the protein nor the DNA change their structures significantly in the nonspecific complex is in agreement with previous observations from the EcoRV and BamHI systems (*21*, *22*).

Because biomolecules are inherently dynamic, analyses of time-averaged structures do not provide a complete description of the mechanisms involved in biomolecular recognition. We have therefore characterized the dynamic properties of the protein at all three stages in order to monitor how the dynamics of the system adjusts along the protein-DNA recognition coordinate. NMR relaxation of backbone amides provides a powerful experimental approach for detecting conformational dynamics of proteins at atomic resolution (*23*). The fast pico- to nanosecond (ps-ns) motions of structured (α helical) regions are restricted and do not vary along the pathway (fig. S2). These motions are different only in the loop between helices II and III (around His^{29}) and the hinge region (residues 52 to 58), where they become quenched upon operator binding. Slower domain motions on the micro- to millisecond time scale (μs-ms) are biologically very important, because they are close to the time scales on which docking, protein folding, allosteric transitions, and product release take place and are therefore associated with functional processes (*23*). Motions on this time scale are identified by the term R_{ex}, which indicates exchange between conformations that sense different chemical environments (SOM Text). Many residues in the free state show prominent micro- to millisecond time-scale motions (Fig. 4, A and B). Binding to the nonspecific sequence considerably increases the R_{ex} term of most residues located in the protein-DNA interface (Fig. 4, A and B) (*24*). Interestingly, most of the residues that exhibit enhanced mobility on the μs to ms time scale are involved in DNA binding in both the nonspecific and the specific complex. Whereas almost all of the residues with enhanced R_{ex} interact with the DNA backbone, some of them adopt alternative conformations in the structural ensemble. For instance, the side chains of the recognition residues Q18 (*25*) and R22 adopt alternative conformations with a total energy very similar to that calculated from the

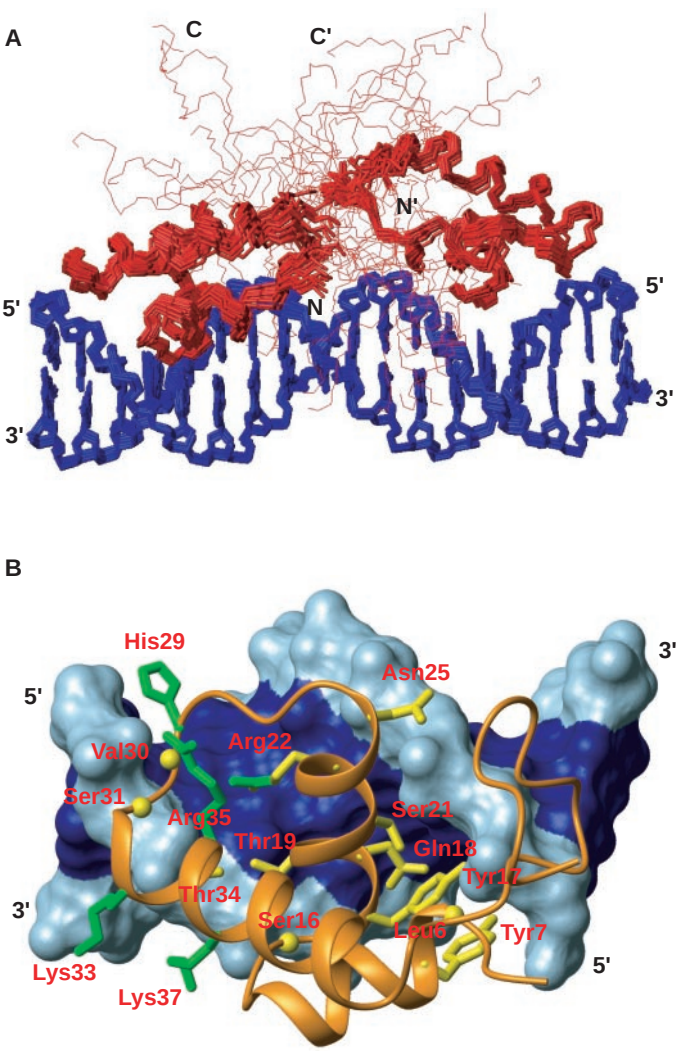

Fig. 1. Structure of the dimeric *lac* DBD complexed to a palindromic nonspecific DNA. (**A**) Ensemble of the final 20 structural conformers. The C terminus (residues 50 to 62) of each of the dimers is unstructured. Protein backbone is depicted in red, whereas the DNA heavy atoms are depicted in blue. The thin red lines depict the unstructured C-terminal region (51 to 62). The final structures have been deposited in the PDB (accession code 1OSL). (**B**) Structure of the left site of the complex. A ribbon diagram of the protein is shown bound to the solvent-accessible surface of the DNA. The major and the minor grooves of the DNA are colored dark blue, and the ribose phosphate backbone, light blue. Residues involved in intermolecular hydrogen bonding and Coulombic interactions are shown in yellow and green, respectively. Backbone amides are indicated with spheres. The structure has been rotated by 80° about the DNA axis relative to (A).

structure determination protocol, with all of them satisfying the experimental constraints. Thus, the remarkable flexibility of the protein-DNA interface may primarily originate from the recognition-helix residues sampling different base pair environments in the nonspecific complex. Because motions on the μs-ms time scale indicate thermally assisted conformational transitions, the high flexibility of the complex may be of advantage for efficient probing of binding sites. The prominent conformational dynamics exhibited by these residues in the free state and the nonspecifically bound state are completely quenched upon binding to the cognate operator, consistent with the selection of a single conformer (Fig. 4, A and B). These results support experimentally the view of nonspecific versus specific binding in which the less-selective complexes are expected to have a wider and more rugged bottom of the binding energy funnel, with low barriers between conformers of the complex (*26*). In contrast, highly specific complexes are relatively rigid, with a steep funnel of conformations leading to the native structure.

It is now well established that the native state ensemble of a protein consists of an equilibrium ensemble of microstates with different ligand binding affinities (*27*). H-D exchange rates of amide protons were measured by NMR for the *lac* DBD at all three stages of the reaction coordinate in order to monitor how the native state ensemble is redistributed along the recognition pathway. In the free state, the protein backbone amides are only marginally protected, suggesting a high population of open microstates (conformations from which exchange can take place) (Fig. 4C) (*28*). Binding to the NOD fragment does not significantly increase the protection factors, thus no redistribution of the protein ensemble takes place (Fig. 4C). In contrast, binding to the specific operators dramatically reduces the population of open conformational states of *lac* DBD (Fig. 4C). In fact, the extent of redistribution is modulated by the nature of the DNA sequence. *Lac* repressor binds most tightly to a left half-site symmetrized sequence (SymL), followed by the natural operators O1 and O3 (*12*) (fig. S1). We show that the extent of population narrowing increases as the DNA binding becomes tighter (Fig. 4C). Therefore, *lac* repressor uses the same binding site to interact with both specific and nonspecific DNA sequences, with the equilibrium population of the protein conformational substates redistributed and shifted depending on the DNA sequence. The data suggest a hierarchical model where the biolog-

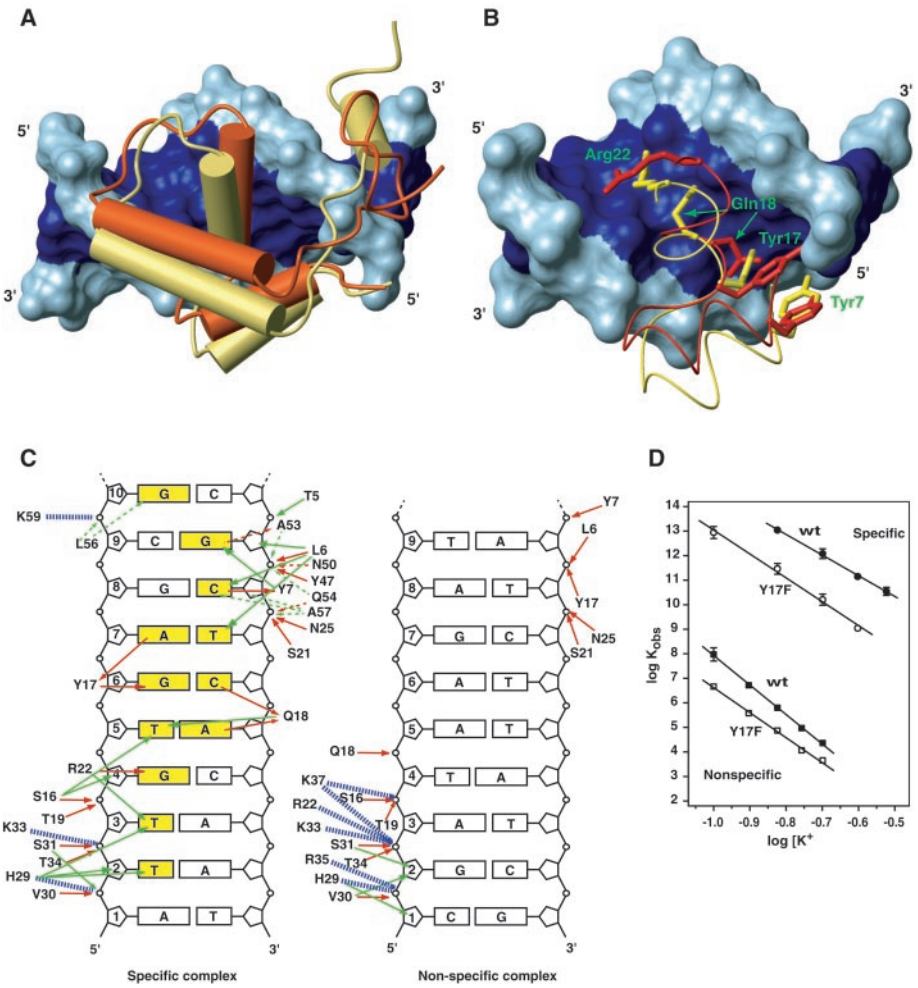

Fig. 2. Comparison of specific with nonspecific binding mode and interactions. In (A) and (B), the left sites of the specific *O1* and nonspecific complexes are overlaid on their DNA backbone so that the protein position with respect to the DNA can be compared. The proteins in the specific and nonspecific complexes are colored yellow and red, respectively. (**A**) The protein in the nonspecific complex is rotated 25°, relative to the DNA, compared to the protein in the specific complex. In (**B**), the four most important residues for conferring specificity are shown and their conformation is compared. (**C**) Schematic summary of the protein-DNA contacts in the specific and nonspecific complex (*25*). The bases that are specifically recognized by *lac* repressor are colored yellow. The solid and dashed lines indicate interactions in the major and minor grooves, respectively. Red, green, and dashed blue lines indicate hydrogen bonding and hydrophobic and electrostatic contacts, respectively. (**D**) Salt concentration dependence of specific and nonspecific binding for wild-type (wt) and Y17F repressors.

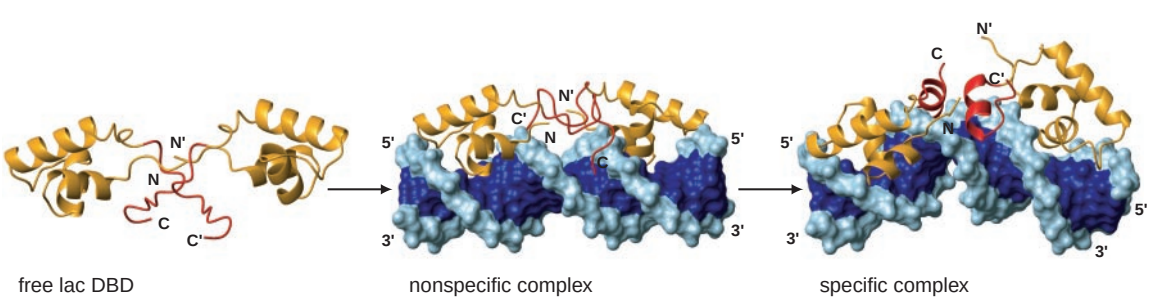

Fig. 3. Structural pathway of protein-DNA recognition. The hinge region (residues 50 to 62), which is colored red, remains unstructured in both the free state and the nonspecific complex, whereas it folds up to an α helix in the specific complex with the natural operator *O1*. In the nonspecific complex, the DNA adopts a canonical B-DNA conformation, whereas in the specific complex it is bent by ~36°.

ical system becomes more ordered going from the nonspecific to the specific complex formation.

It has been long recognized that specific binding is accompanied by a large negative heat capacity change (ΔC_p), whereas a less dramatic change is characteristic of nonspecific binding (*9*). The ΔC_p for *lac* repressor binding to specific operators is ≈–1100 to –1300 cal mol^{-1} K^{-1}, whereas for nonspecific binding it is only ≈–200 cal mol^{-1} K^{-1} (*29*). The variation in ΔC_p of specific compared with nonspecific binding has been generally accounted for by the difference in the solvent-accessible surface area removed from bulk water upon complex formation. However, although formation of the specific complex results in 1400 Å^2 excess of buried surface, its contribution to ΔC_p is the same as for the nonspecific complex (SOM Text). Changes in flexibility and the population of microstates upon binding contribute significantly to ΔC_p (*30*, *31*). The protein-DNA interface in the nonspecific complex is highly flexible on all probed time scales, whereas in the specific complex it is rigid (Fig. 4A and fig. S2). Moreover, considerable narrowing of the microstates population takes place only upon specific and not nonspecific binding (Fig. 4C and SOM Text). Thus, these sources may contribute to the larger value of ΔC_p observed for the specific complex.

The notable charge redistribution, which results in stronger electrostatic interactions in the nonspecific complex, may explain the large and characteristically different salt dependence of specific compared with nonspecific binding (Fig. 2D and SOM Text).

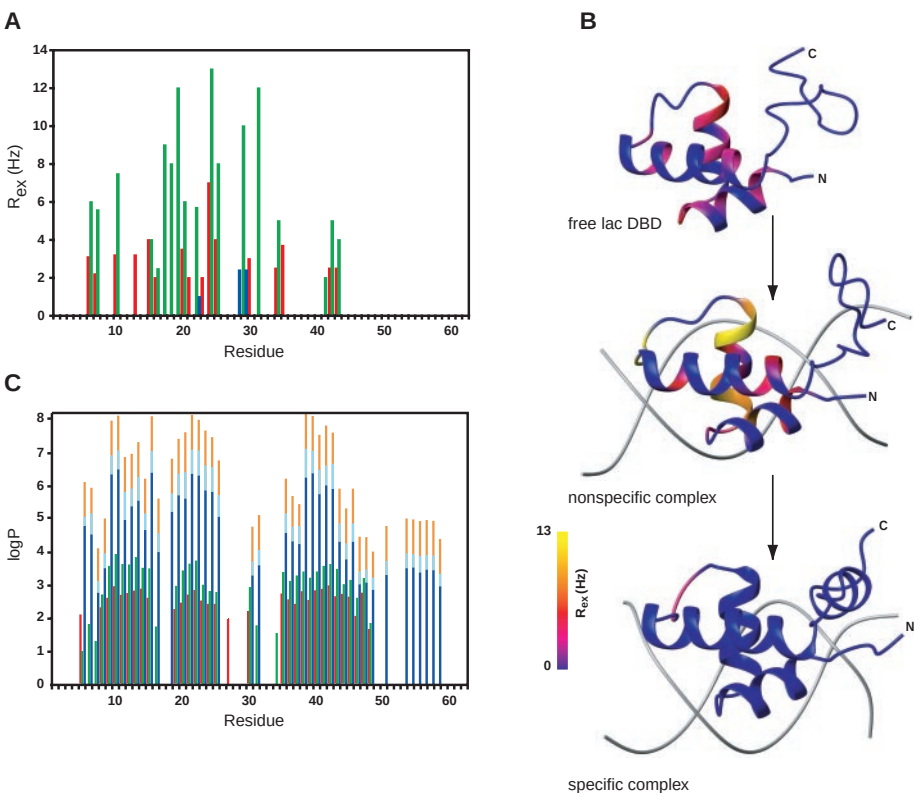

Fig. 4. Dynamic and H-D exchange rates analysis of the dimeric *lac* DBD-DNA interaction. (**A**) Exchange values, R_{ex}, indicating motions on the micro- to millisecond time regime plotted as a function of residue number of the *lac* DBD. Values for the free state, nonspecific, and specific *O1* complex are shown in red, green, and blue, respectively. Only residues with significant R_{ex} values (>1 Hz) are displayed. (**B**) Color-coded representation of the conformational exchange dynamics alteration along the protein-DNA recognition pathway. (**C**) Protection factors of the dimeric *lac* DBD plotted as a function of residue number. Values for the free state, nonspecific, natural operators *O3* and *O1*, and SymL complexes are shown in red, green, blue, cyan, and orange, respectively. Protection factors were calculated from the rate ratio k_{int}/k_{obs} and are displayed as a logarithmic scale.

References and Notes

1. P. H. von Hippel, *Science* **263**, 769 (1994).
2. Y. Kao-Huang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4228 (1977).
3. N. Shimamoto, *J. Biol. Chem.* **274**, 15293 (1999).
4. P. H. von Hippel, O. G. Berg, *J. Biol. Chem.* **264**, 675 (1989).
5. H. Kabata *et al.*, *Science* **262**, 1561 (1993).
6. T. Misteli, *Science* **291**, 843 (2001).
7. D. M. Gowers, S. E. Halford, *EMBO J.* **22**, 1410 (2003).
8. L. Jen-Jacobson, *Biopolymers* **44**, 153 (1997).
9. R. S. Spolar, M. T. Record Jr., *Science* **263**, 777 (1994).
10. C. W. Garvie, C. Wolberger, *Mol. Cell* **8**, 937 (2001).
11. A. D. Riggs, S. Bourgeois, M. Cohn, *J. Mol. Biol.* **53**, 401 (1970).
12. C. G. Kalodimos, G. E. Folkers, R. Boelens, R. Kaptein, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6039 (2001).
13. C. G. Kalodimos *et al.*, *EMBO J.* **21**, 2866 (2002).
14. Each subunit of the dimeric *lac* headpiece (HP) employed for the present studies comprises the first three α helices (residues 1 to 49) plus the hinge region (residues 50 to 62) and a cysteine residue in place of Val^{52} (*lac* HP62-V52C). The V52C mutation was designed so as to link two *lac* headpieces by means of a disulfide bond, and biochemical experiments showed DNA binding parameters of the dimeric headpiece comparable with those of the intact repressor (*12*).
15. The NOD sequence (5′-CGATAAGATATCTTATCG-3′) was designed so as to be entirely different in all positions from the naturally occurring operators. Its binding affinity for the *lac* headpiece is at nonspecific levels as determined by isothermal titration calorimetry. Special care has been taken so that the length of the DNA fragment matches exactly the DNA binding site of *lac* repressor. Longer sequences have been tested, but they result in substantial heterogeneity originating primarily from protein sliding along the DNA.
16. M. C. Mossing, M. T. Record Jr., *J. Mol. Biol.* **186**, 295 (1985).
17. M. Lewis *et al.*, *Science* **271**, 1247 (1996).
18. C. A. Spronk, M. Slijper, J. H. van Boom, R. Kaptein, R. Boelens, *Nat. Struct. Biol.* **3**, 916 (1996).
19. C. M. Falcon, K. S. Matthews, *Biochemistry* **40**, 15650 (2001).
20. Further evidence for the straight B DNA form in the nonspecific complex comes from the characteristic pattern of sequential nuclear Overhauser effects (from one nucleotide to the next) between 1′ (and 2′) sugar protons and H6 or H8 protons on the bases. These patterns continued throughout the whole NOD fragment, both in the free form and in the nonspecific complex. In contrast, this pattern was interrupted at the position of the kink in the specific complex, where the minor groove insertion of the pair of Leu^{56} occurs. Furthermore, the chemical shifts of the central base pairs are identical in the free and bound NOD fragment, consistent with the absence of any structural distortion of the DNA in the nonspecific complex.
21. F. K. Winkler *et al.*, *EMBO J.* **12**, 1781 (1993).
22. H. Viadiu, A. K. Aggarwal, *Mol. Cell* **5**, 889 (2000).
23. M. Akke, *Curr. Opin. Struct. Biol.* **12**, 642 (2002).
24. Most of the residues with high R_{ex} value in the nonspecific complex have the same chemical shift in the free and nonspecific complex state. Moreover, whereas the stability of the nonspecific complex depends greatly on the salt concentration, R_{ex} seems to have an insignificant dependence within a short range of salt concentration used. Therefore, the measured R_{ex} corresponds to motions that take place while the protein is bound on the nonspecific DNA fragment.
25. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
26. B. Ma, S. Kumar, C. J. Tsai, R. Nussinov, *Protein Eng.* **12**, 713 (1999).
27. E. Freire, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10118 (1999).
28. C. G. Kalodimos, R. Boelens, R. Kaptein, *Nat. Struct. Biol.* **9**, 193 (2002).
29. D. E. Frank *et al.*, *J. Mol. Biol.* **267**, 1186 (1997).
30. J. M. Sturtevant, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2236 (1977).
31. M. R. Eftink, A. C. Anusiem, R. L. Biltonen, *Biochemistry* **22**, 3884 (1983).

32. Financial support from the Netherlands Organization for Scientific Research (NWO-CW) is gratefully acknowledged. C.G.K. was supported by a European Union Research Training Network, contract number HPRN CT. 2000 00092. M.L.'s contributions were supported by NIH GM 23467 (T. Record). The atomic coordinates have been deposited in the Protein Data Bank (PDB accession code 1OSL).

Supporting Online Material

www.sciencemag.org/cgi/content/full/305/5682/386/DC1
SOM Text
Figs. S1 to S4
Table S1
References

23 February 2004; accepted 1 June 2004