Toward an Integrated Model of Protein−DNA Recognition as Inferred from NMR Studies on the *Lac* Repressor System

Charalampos G. Kalodimos,[†] Rolf Boelens,[‡] and Robert Kaptein*[,‡]

Department of Chemistry, Rutgers University, Newark, New Jersey 07102, and Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received September 29, 2003

Contents

1. Introduction

Sequence-specific protein−DNA interactions are responsible for the regulation of key biological functions such as replication of the genome, initiation of transcription, and repair of damaged DNA. All of these regulatory pathways are built on the foundation that proteins are able to bind selectively to a particular DNA site in the genome. The most challenging issue for specific protein−DNA recognition is that the target sequence is immersed in a huge molar excess of nonspecific DNA sequences, which also compete for the same protein. The distribution of protein between regulatory and nonspecific sequences determines the occupancy of the target site in the cell and, hence, the transcriptional activity of the corresponding promoter. The problem is much more acute than it is for other protein−ligand interactions, because obviously the protein cannot differentiate between these sites on the basis of their size or shape.[1] For repressor molecules, such as the lactose (*lac*) repressor, the problem of finding the correct site is even more severe, as only a few copies (∼10) of the protein exist in the cell. Thus, to compete with the large excess of nonspecific genomic DNA, these proteins have evolved to bind with a very high specificity ratio (∼10^7).

The mechanisms by which regulatory proteins discern specific target DNA sequences remain a major area of inquiry. The selective binding of a protein to a particular DNA sequence requires the recognition by the protein of a set of steric and chemical features that in total delineate the binding site. Double-stranded DNA has a relatively uniform structure, with a highly negatively charged sugar−phosphate backbone and a core of stacked base pairs whose edges are exposed in the major and minor grooves. Each DNA sequence has a unique chemical and structural "signature" governed by the pattern of the functional groups that are exposed in the DNA grooves and the sequence content; it is this chemical surface that is recognized by proteins. Correlations between available structural and functional data on protein−DNA binding have established that although there are certain amino acid base-pair preferences,[2,3] an amino acid base-pair code that underlies protein−DNA interactions does not exist.[4] Initially, it was thought that sequence specificity comes from hydrogen bonding interactions between the protein and the DNA, due to the requirement for near collinear positioning of donor and acceptor groups. However, the large number of van der Waals interactions found at protein−DNA interfaces imposes steric constraints on the types of side chains and bases that can be accommodated at particular positions, thereby playing a role in sequence selectivity.[5] In addition, water-mediated protein−DNA hydrogen bonds[6,7] and conformational transitions of local protein and/or DNA regions play key roles in stabilizing specific protein−DNA complexes.[8]

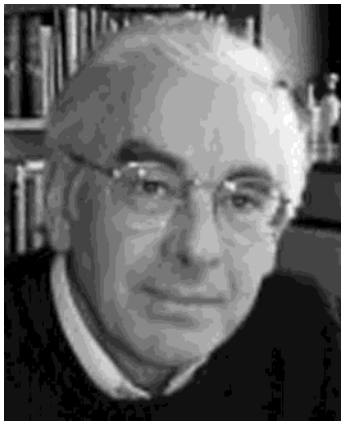
* To whom correspondence should be addressed.
[†] Rutgers University.
[‡] Utrecht University.

10.1021/cr0304065 CCC: $48.50
Published on Web 08/11/2004

Rolf Boelens was born and raised in Groningen, The Netherlands. He received his B.S. and M.S. degrees in chemistry and physical chemistry at the University of Groningen. He completed his Ph.D. on "Ligand Binding to and Electron Transfer in Cytochrome c Oxidase" in the Department of Biochemistry of the University of Amsterdam, working with Bob van Gelder and Ron Wever. Thereafter, he joined the research group of Robert Kaptein, initially as a Postdoctoral Fellow at the University of Gronigen and later as an assistant and associate professor at the Faculty of Chemistry, Utrecht University, The Netherlands. In 1997 he was appointed Professor of Biomolecular NMR Spectroscopy and in 2000 Director of the High-Resolution NMR Facility at Utrecht University. He has worked on EPR of metalloproteins, NMR spectroscopy of DNA binding and ribosomal proteins, and method development in biomolecular NMR spectroscopy. His current interests are biomolecular interactions with an emphasis on transcription and DNA repair. His research has resulted in over 240 publications. He enjoys time outside of work with his wife and two children.

Charalampos (Babis) Kalodimos was born in Volos, Greece, in 1972. He received his B.S. degree in chemistry from University of Ioannina. He started his graduate studies in Institute Curie, Orsay, and received his Ph.D. in Biophysical Chemistry jointly from the University of Ioannina and Institute Curie under the supervision of Professor Ioannis Gerothanassis and Dr. Michel Momenteau. During his studies he received fellowships from the European Molecular Biology Organization and the Federation of European Biochemical Societies to accomplish research projects in solution and solid-state NMR spectroscopy and bioinorganic chemistry at the University of London and the University of Paris 6 & 7. In 1999 he joined the group of Professor Robert Kaptein at Utrecht University as a Postdoctoral Fellow, where he was introduced to the fascinating world of biomolecular NMR spectroscopy. In 2003 he moved to Rutgers University (Newark, NJ), where he is currently Assistant Professor in the Chemistry Department. The main research interests of his group are focused on the elucidation of the molecular basis of protein–protein and protein–DNA recognition and their structure–function relationships by employing NMR spectroscopy and other biophysical methods.

Understanding the mechanisms by which regulatory proteins recognize their target sequences within the DNA genome requires that we also understand the properties of their complexes with nonspecific



Robert Kaptein was born in The Hague, The Netherlands. He received his undergraduate and Ph.D. degrees at Leiden University. His thesis research under the supervision of Prof. L. J. Oosterhoff was on the mechanism of Chemically Induced Dynamic Nuclear Polarization. He has held posts at Shell Research Laboratory, Amsterdam, and at the University of Groningen, and he is now Professor at Utrecht University. He is also Research Director of the Bijvoet Center and Secretary General of the Royal Dutch Academy of Arts and Sciences. He has worked in the area of structural biology of gene regulatory proteins using the methods of NMR spectroscopy.

DNA, since it is the competition between these two types of sequences for the available protein that defines the specificity factor.[9] Given that the specific sequence is a tiny fraction of the total DNA, the initial encounter of a protein with the DNA polymer in vivo will always be with a random nonspecific site. Nonspecific protein−DNA interactions are biologically important because they contribute significantly to the fast in vivo translocation of regulatory proteins.[10] Thus, detailed knowledge of the function, structure, dynamics, kinetics, and energetics of the interaction of a particular protein with both nonspecific and cognate DNA sequences is imperative.

Lactose repressor (*lac*) binding to its operator has long been recognized as the prototypical system for transcription regulation, and their interaction has been the subject of intensive study over the past decades due to their profound biochemical and biotechnological interest. A number of reviews have covered the biochemical and X-ray crystallographic data of the intact repressor bound to its cognate DNA sequences.[11,12] These studies were essential for understanding the topology of the *lac* repressor and the mechanism of induction. However, the crystal structures have failed to yield the subtle features of protein−DNA interactions due to their low resolution and the presence of significant thermal motion in the protein−DNA region.[13−15] In contrast, a very detailed insight into the mechanisms that underlie protein−DNA recognition has been provided by a combination of structural and dynamic NMR studies using the isolated DNA-binding domain of the *lac* repressor as a model system. *Lac* is the only protein−DNA system to date for which the complete recognition pathway (free protein and nonspecific and specific complexes) has been structurally and dynamically characterized. Over the years, the *lac* system has been a proving ground for testing new concepts and developing new techniques. A new methodology that provides a residue-specific view of the association and dissocia-

tion pathway of protein−DNA complexes has also been also developed and tested on this prototypical system. In this review, we summarize how NMR spectroscopy may address unique features of protein−DNA recognition and how the knowledge that has become available through these studies can be combined to give an unprecedented insight into the intricate interactions between proteins and DNA.

2. *Lac* Repressor Interaction with DNA

2.1. *Lac* Repressor Binds DNA in Both the Major and Minor Grooves

The *lac* repressor protein regulates the expression of genes required for lactose transport and metabolism through a process that involves its allosteric interaction with inducer molecules and specific operator DNA sites. The intact repressor is tetrameric, organized as a dimer of dimers, with each subunit (360-residues long) being composed of four functional units: an N-terminal DNA-binding domain (DBD; residues 1−49), a polypeptide linker (residues 50−58), which is referred to as the hinge, a ligand-binding domain (residues 62−333), and a C-terminal tetramerization domain (residues 340−357) (Figure 1).[13] Effective down-regulation of transcription requires that tetrameric *lac* protein binds with one of its dimeric subunits to the primary operator *O1* and

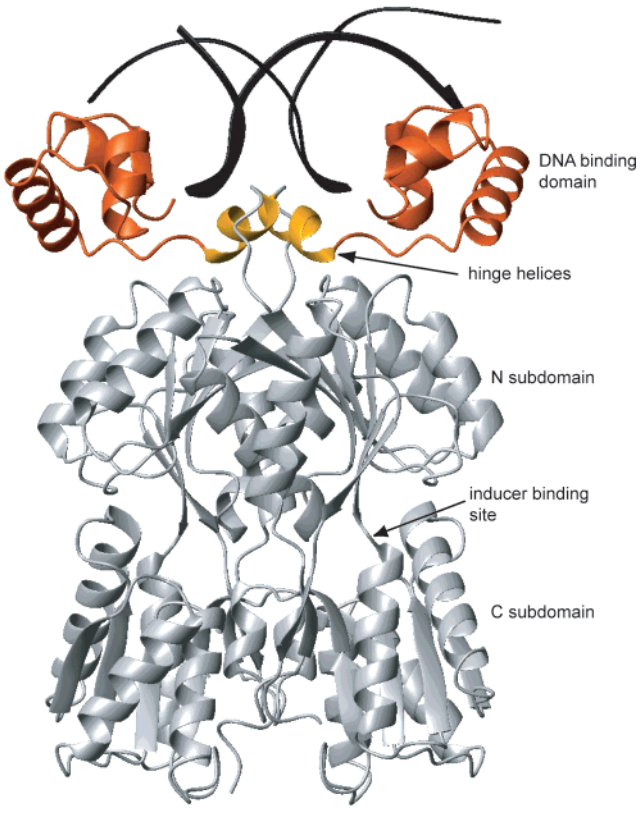


Figure 1. Crystal structure of a dimeric *lac* repressor bound to the SymL DNA operator.[14] The intact repressor is composed of two such units that are assembled through the tetramerization domain (not shown). The DNA-binding domain and the hinge helices are red and orange, respectively. The inducer binding pocket is located at the junction of the N- and C-subdomains. The figure was drawn with MOLMOL[106] using the PDB accession code 1EFA.

simultaneously, using the other dimeric subunit, to one of the two auxiliary operators *O2* and *O3* (Figure 2). Strong interaction with the DNA requires that the

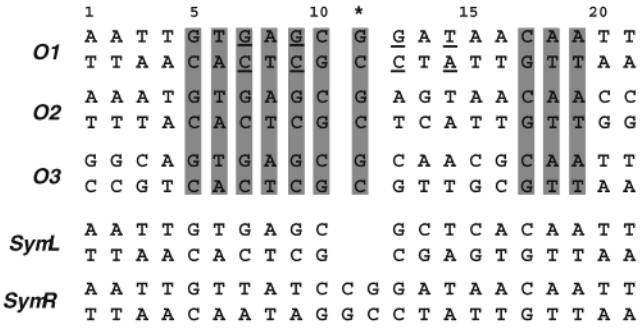


Figure 2. Sequences of the naturally occurring and two symmetric *lac* operators. The three natural operators *O1*, *O2*, and *O3* are aligned with the numbering referred to in the text and figures. The asterisk denotes the central base pair. The bases conserved in all natural operators are highlighted. The two binding sites within each natural operator are asymmetric and are referred to as the left (base pairs 1−10) and right (12−21) sites. In *O1*, the asymmetric regions between the two sites relative to the central base pair are underlined. The SymL operator is a palindrome of the left site of *O1* with the central G:C base pair deleted, and is the sequence used in previous structural studies. The SymR operator is a palindrome of the right site of *O1* including the central base pair.

DBD of *lac* repressor binds as a dimer. Studies of the isolated DBD indicated that this region retains its ability for site-specific binding[16] and thus, in principle, could be used to address the issue of operator recognition. The structure of a short construct of a *lac* headpiece (residues 1−51; HP51) bound to an 14-mer operator (half-site) was the first protein−DNA complex ever solved by NMR spectroscopy.[17] The N-terminal headpiece contains a helix−turn−helix motif (HTH), whose function is to recognize the operator DNA. The HTH motif was first identified as a conserved sequence element in the repressors encoded by the lambdoid phages of *E. coli*.[18] Subsequently, this element has been found in a large variety of DNA-binding proteins, both in prokaryotes and in eukaryotes, and it shows a very high degree of structural conservation.[19] It consists of two short α helices which are separated by a short linker containing a glycine residue. This amino acid, in concert with its neighbors, acts as a flexible hinge allowing the polypeptide chain to make a turn between the two helices so that they can make hydrophobic contacts with each other. The second of the two α helices, referred to as the recognition helix, inserts into the major groove and forms both base and sugar−phosphate backbone contacts. Although the structure of this domain is highly conserved, the orientation of the motif with respect to its DNA binding site is quite variable.

The X-ray structure, first, of the highly homologous *pur* repressor[20] and later of the tetrameric *lac* repressor complexed to a left-site symmetrized operator[13] showed that the polypeptide segment connecting the DBD with the core (residues 50−58) forms an α helix, which is referred to as the hinge helix. However, the low resolution of 4.8 Å of the intact *lac* repressor

structure and the presence of thermal motion in the DBD region hindered the observation of protein−DNA contacts. NMR studies of a 62-residue long headpiece (HP62) construct comprising the hinge region demonstrated that this region is unstructured in the free state, and it folds up to an α helix only when the protein is bound to its operator (Figure 3).[21,22] In contrast, a similar construct of *pur* was

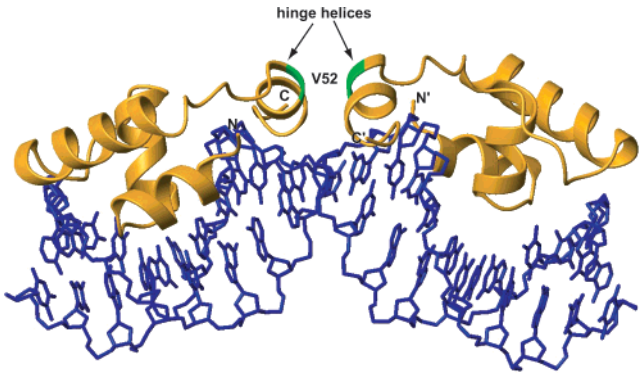


Figure 3. NMR solution structure of wild-type *lac* HP62 bound to the SymL operator (PDB access code 1CJG).[22] The two hinge helices (residues 50−58) bind to the minor groove of the SymL operator and bend significantly the DNA. On the basis of this structure, Val52 (green) was later replaced by a cysteine residue, so that a disulfide bond could link the two subunits, to yield a covalently linked dimeric lac that bound the natural operator *O1* with very high affinity.[35]

shown to be incapable of inducing hinge helix formation upon specific binding to the *purf* operator, suggesting that the core domain of the *pur* repressor is critical for stabilizing the hinge helices.[23] The combined structural studies revealed that the two headpieces bind symmetrically to the operator with their HTH motifs inserted into the major groove. The hinge helices, which run antiparallel, form the interface of the two headpieces and contact the bases of the minor groove in the center of the operator. Intercalation of Leu56 side chains pries apart the central base step, and a 45° kink that bends DNA away from the protein, together with unwinding and base unstacking at the central base step, opens the minor groove sufficiently to enable residues in the hinge helix pair to form direct contacts with bases in the minor groove. The binding mode of *lac* repressor to DNA is essentially the same as that of purine repressor (PurR),[20] and they are notable examples of complexes where the DNA is significantly deformed to accommodate the protein fold. The great majority of protein−DNA complex structures contain DNA that is essentially B-form, with only a moderate degree of bending and deformation. In these cases, it is the surface of the protein that conforms to the DNA structure. While deforming the DNA requires energy, the cost can clearly be compensated by a sufficient number of favorable contacts with the protein.[24,25] In most cases deformation appears to be incidental to complex formation, although, in proteins such as BmrR, the transcription regulator for the multidrug transporter gene *bmr* in *B. subtilis*, the observed DNA distortions are thought to have functional significance for gene regulation.[26]

2.2. Hinge Helices and Allosteric Mechanism

The hinge region is central to the biological function of the *lac* repressor, as it plays a crucial role in the induction mechanism. Inducer binding to the core domain is believed to alter the network of interactions between the core and the DNA-binding domain, thereby destabilizing the binding of the hinge helices to the minor groove of the operator.[13,14] The significance of the hinge region, which is implicated in both the induction mechanism and DNA bending, has motivated extensive biochemical and structural analysis of this flexible segment. For example, mutagenesis data have shown that all residues in the hinge region are highly sensitive to amino acid replacements.[27] Until recently, however, formation of the hinge helices had been unambiguously demonstrated only in the case of *lac* binding to an artificial symmetric operator (SymL; Figure 2). This fragment lacks the central base pair and is a palindrome of the left-operator site. Due to its high affinity for the *lac* repressor, it has been widely used for binding studies.[28] Previous data had indicated that *lac* repressor does not bend the primary operator (*O1*), thus questioning the role of the hinge helices in DNA binding and the validity of the induction model in the wild-type *lac* operon.[29−31] In contrast, a more recent biochemical study provided evidence that hinge helices form when the *lac* repressor is bound to the *O1* operator and that DNA is significantly bent.[32] In support of this conclusion, a low-resolution X-ray structure of a dimeric *lac* repressor bound to the iodinated *O1* sequence showed that for the *O1* operator the bending is similar to that of the symmetric sequences. The NMR spectrum of the *O1* complex with HP62 suffered from unfavorable chemical exchange processes, which resulted in very broad NMR lines (Figure 4A). Thus, a detailed analysis of the hinge helix stability and its difference in binding to the left and the right sites had not been possible until recently.

2.3. Engineering Protein–DNA Affinity

Detailed NMR analysis has shown that the N-terminal domain of the *lac* repressor binds operator DNA in an almost identical way to that of the intact protein,[22] yet with significantly lower affinity. The reason that the isolated headpiece binds DNA weakly (in the μM range) is that binding of two monomers is entropically less favorable than that of a dimer. In fact, the ability to form dimeric or higher-order multimers is of crucial importance for DNA-binding proteins, since it dramatically increases their binding affinity. Extensive genetic studies on the *lac* repressor have also emphasized this point. Mutations at positions that disturb the dimerization interface result in monomeric protein that completely abolishes the ability to repress.[33,34] Therefore, the design and production of engineered DNA-binding domains that could restore the binding affinity of the intact repressor would be of great importance. The low affinity of the isolated headpiece for DNA had also been the main obstacle for studying the structural and dynamic features of *lac* binding to variant operators.

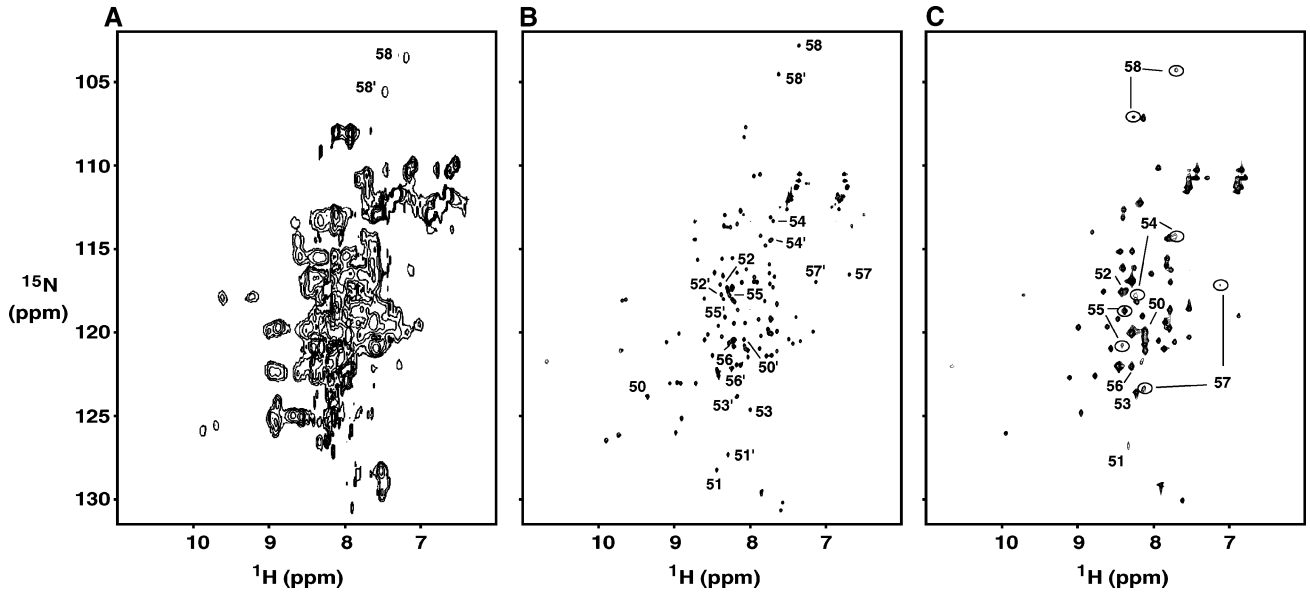


Figure 4. ^{1}H−^{15}N HSQC spectra of (A) the wild-type *lac* HP62 bound to the *O1* operator, (B) the dimeric *lac* HP62−V52C in complex with the natural operator *O1*, and (C) the dimeric *lac* HP62−V52C in complex with the SymR operator. Peaks corresponding to the hinge region (residues 50−58) are labeled. In part B normal numbers indicate residues of the left site while primed numbers indicate residues of the right site. In part C peaks corresponding to the different folding states of hinge region residues are indicated in circles.

Recently, it was demonstrated that a designed V52C mutant with a Cys52−Cys52′ disulfide bond in the hinge region of *lac* HP62 (Figure 3) provides a large increase in the stability of the *lac* HP62−DNA complex.[35] This mutation has been originally introduced in the intact *lac* repressor, and it was designed to generate a protein in which a disulfide bond could be formed between the N-terminal domains.[36] According to the high-resolution NMR structure of *lac* HP62 complexed to the SymL operator (Figure 3), the distance between the Cα atoms of Val-52 and its dimer mate is 5.60 Å, thus matching exactly the optimal distance of disulfide bond formation. Considering also the relative positions of the side chains of the residues, this site is ideal for accommodating a disulfide bond. Remarkably, the disulfide cross-linked (oxidized) HP62−V52C mutant exhibited a (3×10^4)-fold higher DNA-binding affinity as compared with the reduced form, reaching an apparent K_d value of 30 pM.[35] Under the same conditions, the tetrameric *lac* repressor binds DNA with similar affinity, suggesting that rationally designed DBD can be used to restore the binding affinity of the intact proteins. The "chelate effect" has been frequently exploited, and several DBDs have been fused either with a peptide linker or disulfide bonds to yield dimers with increased affinity for DNA. Examples include the λ repressor,[37] the 434 repressor,[38] and the Arc repressor.[39] However, the enhancement of binding strength reported for the HP62−V52C construct has been by far the highest reported in the literature. The high affinity and the relatively small size of this dimeric *lac* headpiece rendered it an ideal system to investigate in detail the structural and dynamic features of *lac* binding to various operator sequences. Indeed, the increase of the stability of the *lac* HP62−*O1* operator complex resulted in the dramatic improvement of the quality of the NMR spectra (Figure 4B).

3. Plasticity in Protein–DNA Recognition

3.1. *Lac* Protein Binding to Its Natural Operator

Pseudo-dyad related sequences are commonly found as the DNA target sites of both prokaryotic and eukaryotic multimeric gene regulatory proteins. Structural data on these complexes would be very interesting, since they may reveal the mechanisms that proteins use to recognize different DNA sequences within the context of a unique operator. However, in NMR studies it has been a common strategy to use symmetric sequences to force identical and symmetric protein−DNA interactions in order to reduce the assignment task and the number of NOESY cross-peaks by 2-fold.[40] Similarly, only a small set of crystal structures of dimeric proteins bound to asymmetric DNA target sites have been obtained, as in most cases altered sequences with idealized symmetry for improved crystallization and diffraction have been used. The recent high-resolution structure of the dimeric *lac* HP62 bound to its natural operator *O1* was the first asymmetric protein−DNA complex solved by NMR spectroscopy (Figure 5).[41]

Until recently, all of the detailed structural studies of the *lac* repressor−operator system have employed a fully symmetric "ideal" operator (SymL operator; Figure 2), which has given the best results in both the crystallographic and NMR studies.[13,14,22] However, the absence of the central G:C base pair in the symmetric operator places the two half-sites out of register with one another with respect to the natural operator. If the operators were B-form DNA, the two half-sites of the SymL operator would be spaced 3.4 Å closer along the DNA axis and rotated by 36° relative to the natural operators. These structural variations, which can be further amplified by the intrinsic asymmetry of the two half-sites in the natural operator, point to the possibility of significant

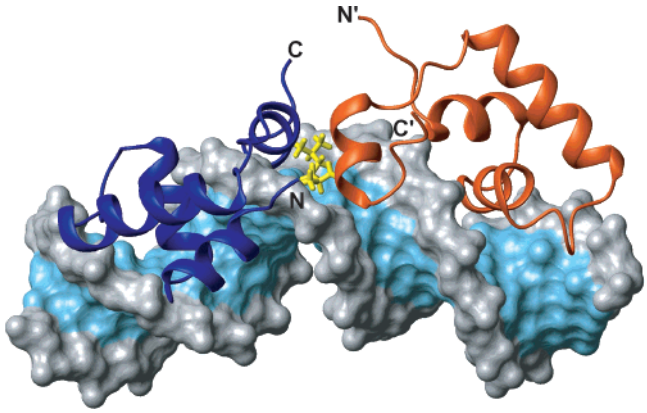


Figure 5. Three-dimensional structure of the dimeric *lac* DBD complexed to its natural *O1* operator (PDB accession code 1L1M).[41] A ribbon diagram of the protein is shown bound to the solvent-accessible surface of the operator. The left and right *lac* headpiece subunits are dark blue and dark orange, respectively. The major and minor grooves of the operator are light blue, and the ribose-phosphate backbone is gray. The side chains of Leu56 (shown in yellow) of both monomers protrude into the minor groove of the *O1* operator and introduce a ~36° kink centered between base pairs 10 and 11.

differences in the way the *lac* repressor binds to the natural operator sequences.[42] The lack of structural data had not allowed assessing the importance of the naturally occurring sequence deviations from symmetry. Within the *lac* operon, each of the three operator sites is pseudopalindromic (Figure 2). The 2-fold symmetry is broken by variations in sequence between the two half-sites and by a central G:C base pair that separates the two half-sites. Relative to the complex with the SymL operator, *lac* repressor could accommodate binding to the natural operator *O1* in two ways: (i) as a rigid protein dimer, which implies that due to the extra central base pair very different sequences are recognized in the left and right half-sites, or (ii) by elongation and twisting of the protein so that the consensus base pairs are recognized in both half-sites, while maintaining the hinge helix interactions (Figure 6A). However, this latter option would entail a considerable change in conformation in one protein monomer with respect to the other. The specific mode of *lac* repressor binding to its natural operator had remained a mystery, and most of the data that had been reported over the past 15 years were controversial.[43−46] This controversy has been further amplified in view of the outcome of recent structural studies. A low-resolution crystal structure of a dimeric *lac* repressor bound to the *O1* operator suggested that the two DBDs bind to DNA in the same orientation by shifting the right subunit as a rigid body by one base pair toward the center of the operator.[15] In contrast, a high-resolution NMR structure of the dimeric HP62 construct of the *lac* repressor bound to the *O1* operator (Figure 5) revealed a dramatically asymmetric global positioning of the dimer on the operator, which results in a different pattern of specific contacts between the two sites.[41] Below we highlight the most important features of the protein−DNA contacts of the complex (which are discernible only in the NMR structure), since the set of recognition mechanisms involved in

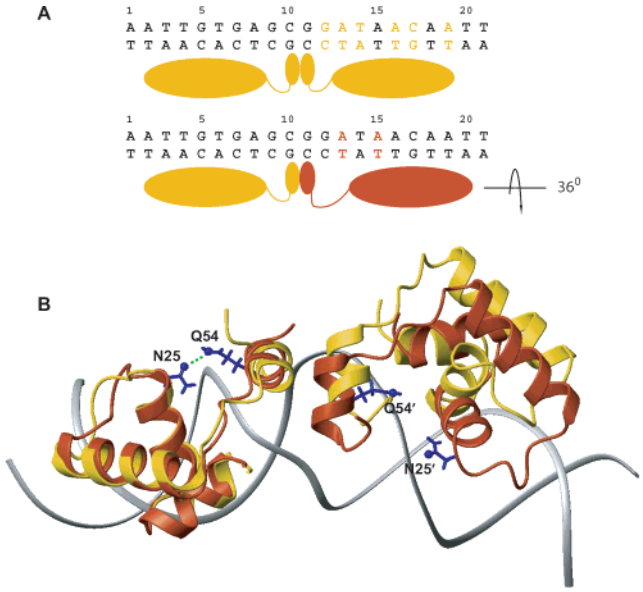


Figure 6. (A) Possible modes of *lac* repressor binding to the natural operator *O1*. According to the first, lac repressor's DBDs bind as rigid bodies, and so, by virtue of the intrinsic asymmetry of the DNA sequence, the right headpiece encounters an almost entirely different sequence relative to the left one (shown in yellow). If the right DBD moves one base pair further away from the center (second model) then it is juxtaposed to a very similar sequence (red). (B) Superposition of the dimeric *lac* HP62 structure bound to SymL (yellow) and the natural operator *O1* (red). The hinge helix in the right site moves by ~3.4 Å closer to the center of the sequence, and the recognition helix rotates by 48° relative to the left one. In the left site of both structures, the hinge helix is stabilized through a hydrogen bond (green dashed line) between Gln54 and Asn25. The extension of the loop linking the third helix with the hinge helix in the right site of the *O1* complex results in the disruption of this critical contact, thereby diminishing the stability of the right hinge helix.

the *lac* repressor−operator system present a nice example of the adaptability that both proteins and DNA exhibit in the context of their mutual interaction. Then, we come back to the discrepancy between the crystal and solution structures.

3.2. Protein Adaptability

The recognition helix of both headpieces makes extensive contacts to the major groove of the natural operator while the hinge helices penetrate, as expected, into the minor groove between base pairs 10 and 11, thereby introducing a kink of ~36° in the DNA (Figure 5). The structure of the left site of the complex is similar to the structure of the *lac* headpiece bound to the SymL operator (Figure 3). However, due to the increased stability of the complex afforded by the introduction of the disulfide bond in the protein interface, a larger number of intermolecular NOEs were collected, which in combination with hydrogen bond constraints yielded a much better defined structure. Extensive chemical shift and NOE analysis comparison of the wild-type HP62 and the covalent dimer HP62 complexed to the SymL operator demonstrated that the covalent link does not introduce any additional effect. Remarkably, the right protein subunit in the *O1* operator complex

adopts an alternative conformation in order to specifically recognize the right half-site of the natural operator (Figure 6B). Although the hinge helix binds to the minor groove between base pairs 10 and 11, the three-helical domain shifts by one base pair further away from the center. To form an optimum interface with the right half-site of the natural operator, the right *lac* headpiece undergoes a 48° rotation relative to the left monomeric site (Figure 6B). This is the result of the shift by one base pair (the helix twist of B-DNA is 36°) and an additional 12° rotation of the recognition helix apparently needed for maximizing the interaction with the right half-site sequence. Therefore, the two protein subunits align in a very different way with respect to the center of the operator in order to achieve optimum juxtaposition of the protein−DNA interface in the two sites.

The conformation adopted by the right *lac* headpiece results in an extension of the loop linking the third and the hinge helix in the right subunit (Figure 6B). This has some important implications regarding the stability of the complex. In the left subunit, Asn25 makes a key hydrogen bond via its side chain CO to the side chain NH_2 of the hinge helix residue Gln54, thus providing a critical link between the core of the *lac* headpiece and the hinge helix (Figure 6B). This link is expected to contribute significantly to the stability of the hinge helix and thus of the dimer interface, and is also present in the highly homologous *pur* repressor.[20] This H-bond is also present in more than 50% of the structural conformers of the wt-HP62−SymL structure, despite the fact that this structure is not as well defined as the dimeric HP62 structure. In the right *lac* headpiece subunit, however, these two groups are far apart, due to the structural rearrangement, rendering impossible the formation of the corresponding hydrogen bond (Figure 6B). This may explain the lower stability of the hinge helices in the complex with the natural operator as compared to that with the SymL operator (see below). Therefore, when the *lac* repressor binds to right symmetrized DNA fragments, the hinge helices would be expected to be even more unstable. To test this hypothesis, a right-site symmetrized full operator, SymR, was used (Figure 2). Interestingly, in the complex of the *lac* headpiece with the SymR operator the hinge helices are not well ordered and exist at equilibrium between α-helical and random coil conformations (Figure 4C).[35] The alternative conformations that the *lac* repressor adopts when bound to different sequences may be the structural origin of the different allosteric response of the *lac* repressor to inducer molecules.[47]

The extensive NMR studies of the *lac* DBD interaction with its operators over the years have provided a wealth of information. Ultimately, all of these efforts culminated in the structure determination of the *O1* operator complex. This structure is the best defined one so far, due to the large number of intermolecular constraints, and it allows for a very detailed comparison of the protein−DNA contacts in both sites of the natural operator. Below, we highlight its most important features that provide an atomic view of the recognition code. All protein−DNA contacts are summarized in Figure 7. Despite the alternative conformation assumed by the two subunits, the *lac* headpiece makes extensive hydrogen bonding contacts to the sugar phosphate in both sites through the same set of residues. Interestingly, Leu6 and Tyr7, the first residues of the HTH, participate in numerous sequence-specific contacts in the left site, whereas these intermolecular contacts are not conserved in the right site of the complex, where the corresponding residues are involved only in nonspecific contacts. Therefore, the first helix residues do not confer any specificity to the recognition of the right-site sequence. Tyr17 and Gln18, both key residues for specificity, make extensive hydrogen bonds to the bases in both sites. The side chain of Arg22 is poised to interact favorably with the base of Gua5, which is in agreement with mutational data.[48] The His29−Ser31 loop makes identical contacts in both sites with the highly conserved Thy3−Thy4 base step. Both hinge helices protrude into the minor groove between base pairs 10 and 11 and introduce a significant kink in the operator with a mechanism that is identical to the symmetric sequences. Asn50 and Gln54 hydrogen bond to the DNA backbone, whereas Ala53, Leu56, and Ala57 are involved in extensive hydrophobic interactions with the partner hinge helix and the bases at the center of the operator.

The 2-fold symmetry of the natural operator is broken at two sites; the G:C base pairs at positions 7 and 9 become A:T at the right site (at positions 13 and 15) (Figure 2). The G:C base pair at position 9 is recognized through specific contacts by Leu6 and Tyr7; these contacts are missing in the right site. Interestingly, the A:T base pair at position 13 is recognized by Tyr17, the side chain of which hydrogen bonds to the N7 atom of Ade13. This contact is achieved by an alteration of the side chain conformation of Tyr7 and Tyr17 in the right site, in which an aromatic-ring-stacking interaction reorients the two tyrosines relative to the left site (Figure 8). In the left half-site, Thy8 is specifically contacted by the methyl groups of Leu6. In the right site, however, there is a complete reorganization: the symmetry related Thy14 is contacted at its methyl group by Tyr17, Gln18, and Ser21 in a specific manner. This side chain rearrangement, in conjunction with an additional rotation of 12° of the recognition helix relative to the left one, allows Gln18 to recognize three different bases, instead of two in the left site (Figure 7). Therefore, despite the global shift of the right HTH domain by one base pair further away from the center, the side chains of Tyr17 and Gln18 adjust locally and move toward the center of the operator, thereby forming entirely different contacts to the operator compared to the case of the left site. Overall, there are surprisingly extensive structural differences between the left and right half-sites with respect to protein−DNA interactions, including a shift of the right *lac* headpiece by one base pair and a rotation of 48° relative to the left one. In most of the asymmetric protein−DNA structures that have been reported so far by X-ray crystallography, the

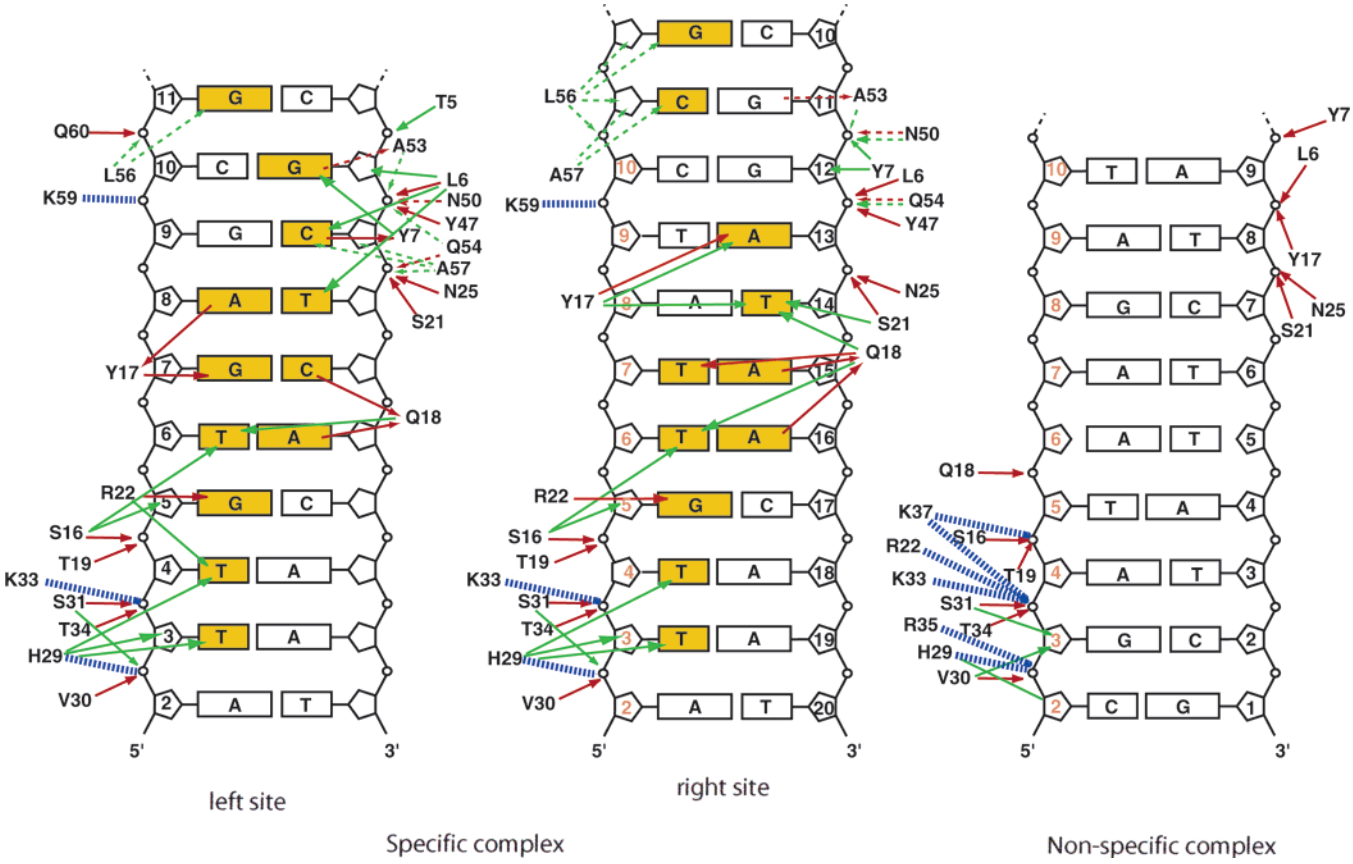


Figure 7. Comparison of the protein−DNA contacts in the two half-sites of the dimeric *lac* HP62−*O1* operator complex and the nonspecific complex. The bases that are specifically recognized by the *lac* repressor are yellow. The solid and dashed lines indicate interactions in the major and minor grooves, respectively. Red, green, and dashed blue lines indicate hydrogen bonding, hydrophobic, and electrostatic contacts, respectively. In the right site of the specific complex and in the nonspecific complex, the numbering displayed in orange denotes the symmetry related base pairs in the left half-site of the specific complex.

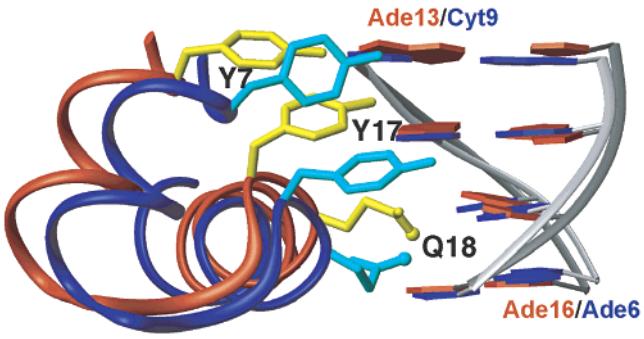


Figure 8. Specific recognition of the two sites of the natural operator *O1* requires also significant rearrangement of the side chain conformation (the three most important residues (Y7, Y17, and Q18) are depicted) as well as a sequence-dependent deformation of the operator. Dark blue and dark orange refer to the left and right sites, respectively. Reprinted with permission from ref 42. Copyright 2002 Oxford University Press.

protein subunits align roughly symmetrically on the DNA half-sites, despite local structural adjustments.[49−52] The *lac* repressor's DBD adopts a novel mode for recognizing its pseudosymmetric natural operator sequence, which is in fact a combination of the strategies employed by a variety of DNA-binding proteins.

As was mentioned above, a low-resolution crystal structure of a dimeric *lac* repressor bound to the *O1* operator suggested that both DBDs bind with the same orientation with the right-side DBD shifted toward the center.[15] In both NMR and X-ray structures a similar bend in the DNA and a close interaction between the hinge helices inserted between base pairs 10 and 11 are observed. However, the structures are different in the position of the three-helical DBD on the right side such that in the X-ray structure it is juxtaposed to an almost completely noncognate DNA sequence. What is the origin of this discrepancy? First, the interaction of the disulfide bond cannot be the reason, since the hinge helices are packed identically in the NMR and X-ray structures. The presence of the repressor core in the X-ray structure could in principle be the origin of the differences. However, we believe that this is not the case. The most important argument is that the NMR structure is in much better agreement with biochemical data accumulated already in the 1970s. Thus, protection of purine methylation by bound repressor extends to base pair 3 on the left side and to base pair 19 on the right side.[53] Repressor−operator cross-linking by UV irradiation of operators in which thymines are substituted by bromouridine extended to thymine at position 22, even further away from the center on the right side than on the left side (Thy1).[54] Finally, base substitution experiments also showed an effect on binding affinity more distant from the center for the right operator side (Thy25 vs Thy4). Thus, the combined evidence of the chemical

modification experiments is not consistent with a shift of the DBD toward the center of the operator. It should be mentioned that due to the low, 4.0 Å, resolution of the crystal structure, many details cannot be observed. For instance, the loop consisting of residues 26−31, crucial for interactions distant from the center, is not seen clearly in the electron density. Furthermore, the only consensus contact between the core of the *lac* repressor and its DBD is the Arg118′−Ile48.[14] Ile48 is located in a flexible loop (as evidenced by relaxation data; see below) between the third and the hinge helix, and it is very doubtful that only this contact could force the headpiece to recognize a sequence in the right site that is entirely different from that in the left site. If this were the case, then deletion of base pair 12 would be expected to have a positive contribution to the binding affinity, as the right DBD would encounter a sequence very similar to that encountered by the left one (Figure 6A). However, this deletion is deleterious for *lac* repressor binding.[55] The G−C base pair in positions 5 (left site) and 17 (right site) is the only absolutely conserved base pair contacted by the recognition helix in all of the six half-sites in the natural operator sequences (Figure 2), and their mutation severely compromises *lac* repressor−operator interaction.[56] If the right headpiece moved one base pair toward the center, then the critical Arg22−G17 contact would be out of register. This same would be true for the conserved base pairs at positions 18 and 19, which would not be reached by the headpiece residues. Overall, on the basis of structural and biochemical data, we believe that the intact repressor recognizes the *O1* operator in the same way as the dimeric headpiece construct seen in the NMR structure.

3.3. DNA Deformability

Understanding the role DNA plays in facilitating the association of DNA-binding proteins is necessary for understanding how sequence specificity is accomplished, since DNA responds to protein binding through sequence-dependent kinking and intercalation. This deformability is essential at both the global and local levels, serving as a potential long-range signal for molecular recognition as well as accommodating the local distortions of the double helix induced by tight binding.[57] The latter is usually referred to as the indirect readout of DNA residues.[58] DNA distortion also affects the thermodynamics of protein−DNA recognition. A recent analysis revealed that formation of complexes with relatively undistorted DNA is driven by a favorable enthalpy change, while the entropy change is unfavorable; the reverse phenomenon was observed for complexes with highly distorted DNA.[59] The observed parameters of the *O1* operator bound to the *lac* repressor DBD (base-pair roll, twist, and major and minor groove width) are summarized in Figure 9. The kink is reflected in increases in the roll and twist angles of the central base pairs and in deviations of the major and minor groove widths and depths, compared with standard B-DNA values. All parameters in the left half-site of the operator sequence are normal for undistorted B-DNA. However, the base pairs of the triplet A:T:A

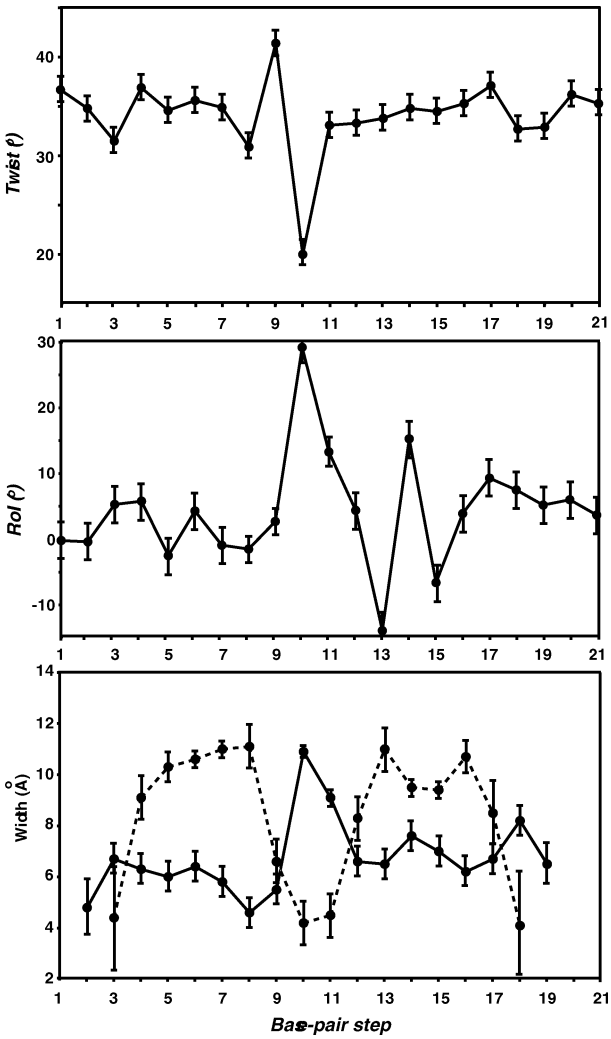


Figure 9. Averaged helical parameters of the *lac* operator *O1* in complex with the dimeric *lac* HP62. The twist, roll, and width of the major (dashed line) and minor (solid line) grooves are shown. Reprinted with permission from ref 42. Copyright 2002 Oxford University Press.

in the right half-site show significant local deformations with base-pair roll that deviate significantly from the averaged values of B-DNA. In addition, the major groove in this region is narrower by ∼2 Å compared to that in its symmetry related region in the left half-site. The conformation of the A:T:A triplet differs significantly from that in its symmetry related region in the left site, where a G:A:G triplet exists, and it imparts asymmetry by providing unique contacts to the HTH domain of the *lac* headpiece. The specific contacts to base pairs provided by Tyr7, Tyr17, and Gln18 are very different between the two-half-sites. Apparently, the base pairs at positions 13, 14, and 15 should deform significantly for optimum interaction with the side chains of the residues located in the recognition helix. Therefore, appropriate recognition of the right half-site requires that the *lac* repressor assumes an alternative conformation in conjunction with the sequence-dependent ability of the half-site to adopt the required conformation upon repressor binding. These results contribute to an evolving view of the importance of the sequence-dependent conformational flexibility of the DNA on protein recognition and affinity.[60,61]

4. Structure and Dynamics Adaptation along the Protein–DNA Recognition Pathway

4.1. Importance of Protein–Nonspecific DNA Interactions

The critical missing piece toward constructing the puzzle of the physicochemical mechanisms that underlie the protein−DNA recognition process is the knowledge of the nature of interaction of the regulatory proteins with nonspecific DNA. Because in the cell nonspecific sequences compete with the target operator for the available regulatory protein, establishing the origin of the recognition requires intimate comparison of protein binding to cognate and noncognate sequences. Protein−nonspecific DNA interactions may also play an important role in the in vivo translocation of DNA-binding proteins during the search for their operator site.[62,63] Indeed, it has been demonstrated that proteins can find their DNA target sites at much faster than diffusion-controlled rates. For example, the *lac* repressor binds its operators 10^2-10^3 times faster than the rate estimated for a 3D diffusion-controlled reaction.[64] A two-step mechanism has been proposed to describe this process. In the first step, a complex with a nonspecific site is formed via a diffusion-controlled reaction. Once bound to DNA, the protein searches for operator sites by a series of intramolecular steps that appear to include sliding (1D diffusion along the DNA backbone) and intersegment transfer (formation of a ternary complex between two DNA sites with a single repressor tetramer) with release to another site within the same DNA molecule. These mechanisms contribute to the high affinity of this regulatory protein for its cognate site by generating a rapid association rate constant for repressor binding to the operator sequence.[63−67] Many site-specific DNA binding proteins make use of linear diffusion along a DNA molecule. In some cases, a strong correlation between the ability of mutant proteins to undergo linear diffusion and their biological function in vivo has been established.[68] RNA polymerase and a small repressor with a HTH motif were directly visualized to slide on DNA by single-molecule dynamics.[62,65] Almost all proteins that interact with specific sites bind nonspecific DNA sequences with appreciable affinity.

As it is now well founded that nonspecific interaction is an important step in the process of sequence-specific recognition and binding, it becomes imperative to describe the structural, dynamic, and thermodynamic response of a protein binding to both nonspecific and cognate operator sequences. Many studies have revealed that sequence-specific binding is coupled to extensive conformational changes in both protein and DNA components.[8,69] How does binding to nonspecific DNA, which precedes the specific binding, alter the structural and dynamic features of the two partners? How does the complex switch from the nonspecific to the specific mode during the exploration process? The question of how binding and specificity are coupled to these conformational changes and to the dynamic processes that occur during these interactions is a very important one.

4.2. Structure of the Dimeric *Lac* Headpiece Complexed to a Nonspecific DNA Sequence

The apparent lack of structural and especially dynamical information on the interactions of sequence-specific DNA binding domains with nonspecific DNA sequences is mainly due to difficulties inherent in studying complexes in which the protein can bind in multiple positions on a DNA fragment. The combined use of the dimeric *lac* headpiece-62 (HP62−V52C) mutant with an 18-base-pair long nonspecific fragment proved to be very successful, since a stable and unique complex was formed.[70] The structural analysis of the nonspecific complex (Figure 10), coupled to the

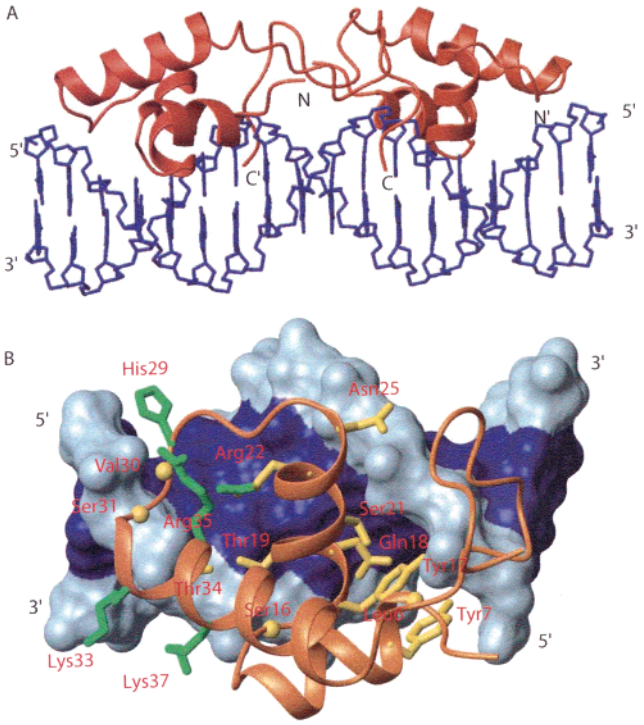


Figure 10. Structure of the dimeric *lac* HP62 complexed to nonspecific DNA. (A) The lowest-energy structural conformer. The C-terminus (residues 50−62) of each of the dimers is unstructured. The protein backbone is depicted in red whereas the DNA heavy atoms are depicted in blue. (B) Schematic diagram of the 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 dark blue, and the ribose phosphate backbone is light blue. Residues involved in intermolecular hydrogen bonding and Coulombic interactions are shown in yellow and green, respectively. Backbone amides are indicated with spheres.

previous studies of the *lac* DBD in the free state and complexed to its cognate sequences, completed the structural characterization of the whole protein−DNA recognition process (Figure 11).

The most remarkable feature of the nonspecific complex structure is that the protein rearranges its backbone and side-chain conformations, as compared to the specific complex, in such a way so as to form an extensive electrostatic and hydrogen bond network with the DNA phosphate backbone (Figure 11). To accomplish this, the protein tilts by ~25° its position relative to the DNA, as compared to the specific complex, resulting in a dramatic alteration

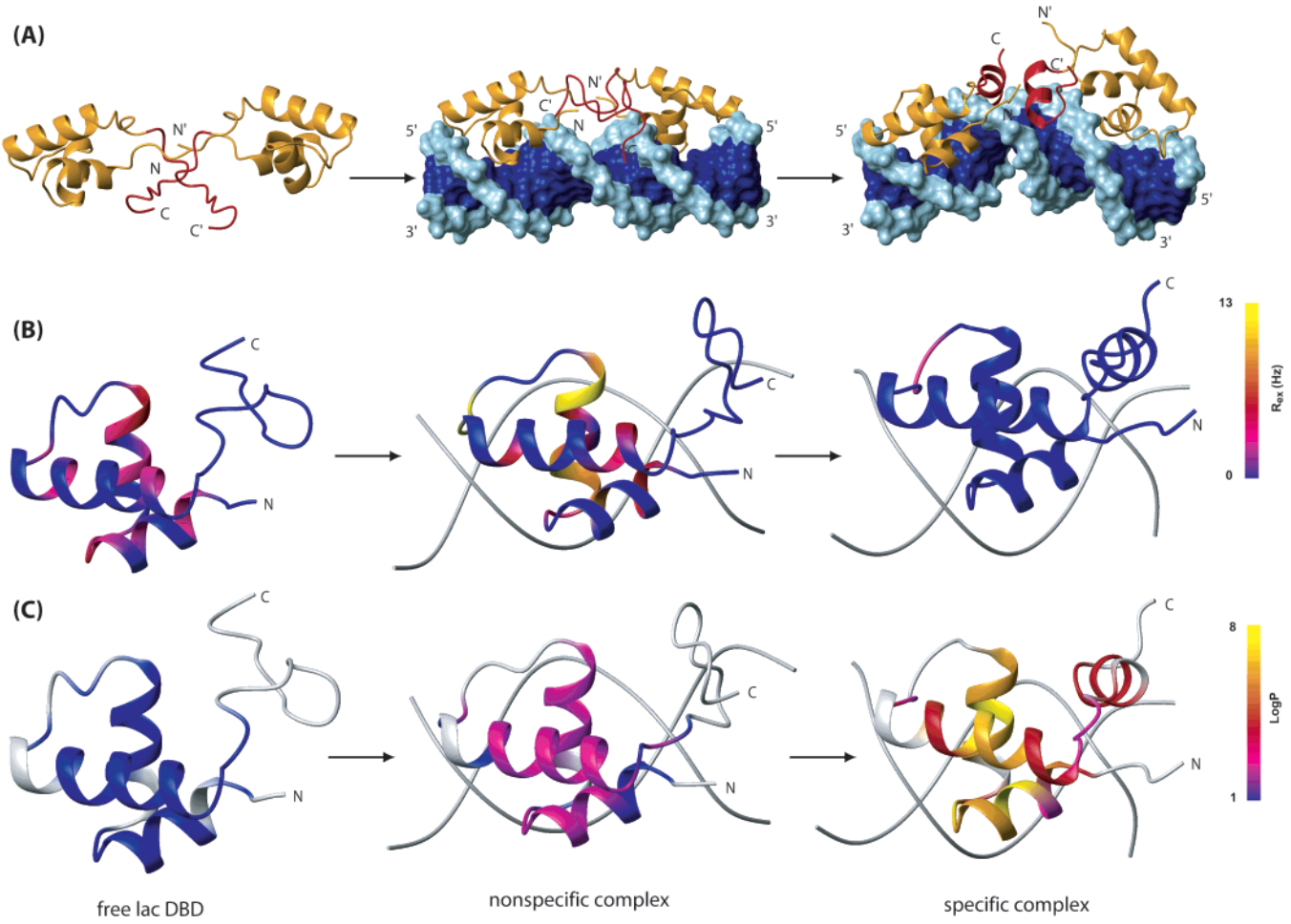


Figure 11. Structural, dynamic, and hydrogen exchange characterization of the pathway of protein−DNA recognition. (A) The hinge region (residues 50−62), which is red, remains unstructured in both the free state and the nonspecific complex, whereas it folds up to an α-helix in the specific complex. In the nonspecific complex the DNA adopts the canonical B-DNA conformation, whereas in the specific complex it is bent by ∼36°. (B) Color-coded representation of the conformational exchange dynamics alteration along the protein−DNA recognition pathway. Exchange values R_{ex} indicate motions in the μs-ms time regime. (C) Alteration of protection factors (*P*) of the dimeric *lac* HP62.

of the protein−DNA contacts (Figures 7 and 12). As was described above, Tyr7, Tyr17, Gln18, and Arg22 are primarily the residues that confer specificity through interactions with the base pairs in the major groove of the specific operators. In contrast, in the nonspecific complex, the side chains of these residues shift and twist so as to participate to hydrogen bond and/or electrostatic interactions with the phosphates. The rearrangement of the protein−DNA surface causes all of the specific base pair interactions that are present in the specific complex to vanish in the nonspecific one throughout the recognition sequence (Figure 7). Apparently, when the protein is located at nonspecific sites, which contain several "incorrect" base pairs, it switches to the nonspecific mode that is totally electrostatic, as was earlier suggested on the basis of thermodynamic analysis.[71] In the specific complex Arg22 interacts uniquely and directly with the bases in the major groove and lies more than 6 Å apart from the phosphates. Compared to the case of the specific complex, two additional charged residues, Arg35 and Lys37, approach the DNA phosphates in the nonspecific complex within a Coulombic interaction distance, thus enhancing significantly its electrostatic character (Figure 7). Numerous protein side-chain and backbone contacts to the sugar phosphates further stabilize the nonspecific complex. Interestingly, all these interactions are preserved in the specific one. The persistence of these intermolecular contacts may be energetically important for a smooth switch from the nonspecific to the specific binding mode once the cognate sequence has been encountered during the exploration process. Therefore, the structural data revealed that the protein presents a scaffold for DNA binding that can be fine-tuned for either nonspecific or specific binding.

In the inducer-bound state, the *lac* protein relieves repression as a result of a dramatic drop of its DNA-binding affinity at nonspecific levels. Inducer binding within the C-terminal oligomerization domain of the repressor (Figure 1) affects the DNA-binding properties of the N-terminal domain located more than 40 Å away. The hinge region (residues 50−58) is postulated to communicate the allosteric signal. Cross-linking the two hinge helices by means of a disulfide bond renders the system noninducible, thus highlighting the role of this polypeptide segment in the induction mechanism.[36] The available structural data demonstrated that, in the free state, the hinge region is disordered and it forms an α-helix when bound to

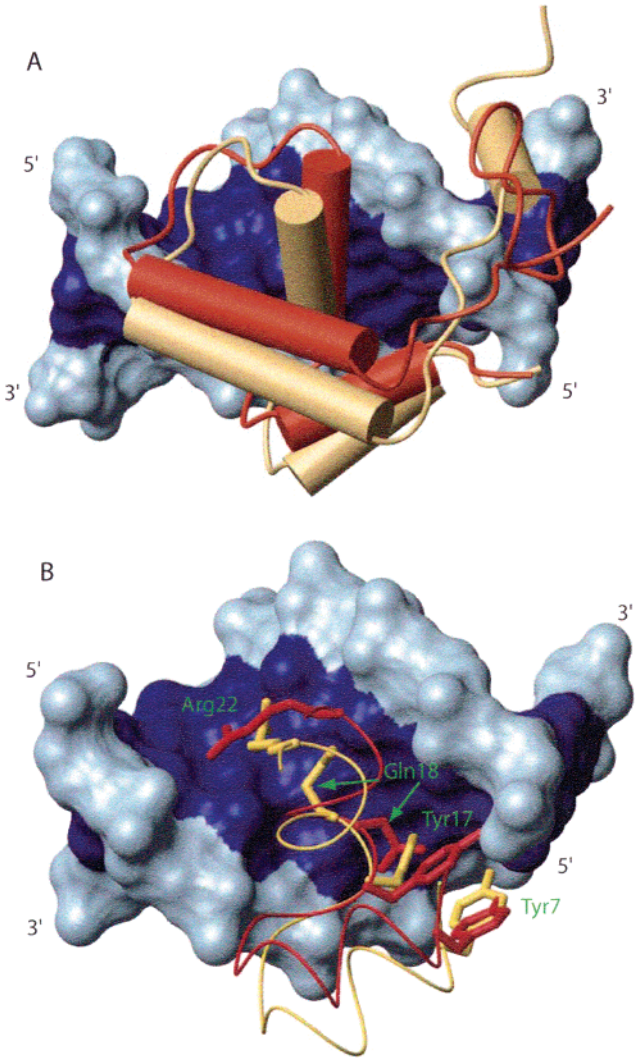


Figure 12. Comparison of specific versus nonspecific binding modes and interactions. The left sites of the specific and nonspecific complexes are overlaid on their DNA backbone so that the protein positions with respect to the DNA can be compared. The protein in the specific and nonspecific complexes is yellow and red, respectively. Part A depicts the 25° rotation on the DNA of the protein in the nonspecific complex. In part B, the four most important residues for conferring specificity are shown and their conformations are compared.

specific DNA (Figure 11). Significantly, binding of the *lac* DBD to the nonspecific sequence was shown to be incapable of inducing α-helix formation, further emphasizing the importance of this region both for specificity and as a structural switch between nonspecific and specific binding modes. This may have a vital implication for the mechanism of target site location, because the *lac* repressor while being bound nonspecifically to DNA scans several hundreds of base pairs by linear diffusion, searching for its recognition site.[64] As a direct consequence of the nonfolding of the hinge helices, no minor groove contacts are present and thus the central kink of ~36° observed in the operator in the specific complex is absent in the nonspecific one, where DNA remains in the canonical B-form (Figure 10). Therefore, a key observation extracted from that study is that both protein and DNA change their structures very little as they bind nonspecifically to each other. This indicates that protein folding and/or DNA distortion is not a significant feature of the nonspecific DNA recognition. These results are in agreement with earlier observations from EcoRV and BamHI systems.[72,73] It has been suggested, on the basis of thermodynamic observations, that the protein−DNA interface in nonspecific complexes retains a significant portion of its hydration. Indeed, three residues (Tyr7, Gln18, and Arg22) were found to form water-bridged hydrogen bonds with bases in the major groove of the nonspecific operator (Figure 13). Due to the major side-chain rearrangement, a cavity is formed in the protein−DNA interface in the major groove that can accommodate water molecules.

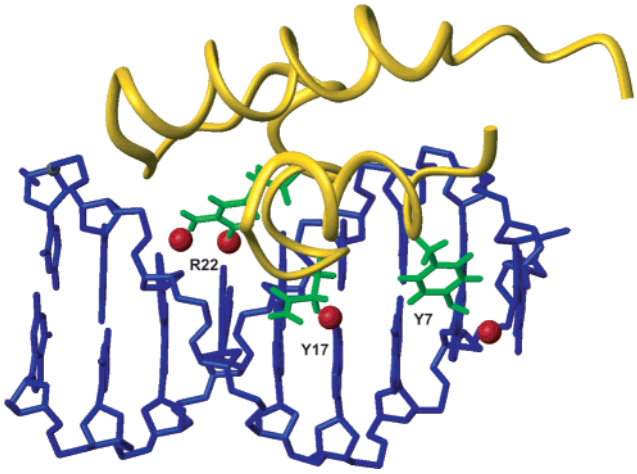


Figure 13. Schematic diagram of the structure of the left site of the dimeric *lac* HP62 complexed to the nonspecific operator. The red spheres represent the water molecules that were consistently seen (in more than 70% of conformers) to mediate interactions of the side chains of residues Arg22 and Gln18 with the bases of the major groove and of Tyr7 with the phosphate backbone.

4.3. Dynamics Is a Central Feature of Protein–DNA Recognition

Because biomolecules are inherently dynamic, analyses of time-averaged structures do not provide a complete description of the mechanisms involved in biomolecular recognition. Protein flexibility and the dynamics of intermolecular interfaces can regulate binding affinity and specificity in molecular recognition, and they can also have a profound effect on determining the thermodynamics and kinetics of the binding process.[74,75] Now it is well-known that proteins have various motions covering a wide range of amplitudes and time scales ranging from picoseconds to hours. These movements may correlate with protein function; however, the relative contributions of the various motions to function are yet to be known. The first step toward understanding the interplay between dynamics and function is to recognize the range and amplitudes of motions present in the system under investigation. In this respect NMR offers a great advantage over all the other techniques and enables detailed characterization of sequence-specific local and global dynamical properties of proteins and their complexes in aqueous solution.[76,77]

The fluctuating local magnetic fields, which cause spin relaxation, arise as a consequence of intramolecular motions and rotational tumbling of the entire molecule. A suite of NMR techniques is now available that allows one to investigate and distinguish intramolecular motions in different "time windows"; in solution NMR these windows are commonly divided into the picosecond to nanosecond (ps-ns), microsecond to millisecond (μs-ms), and second time scales. By measuring the longitudinal relaxation time (T_1), the transverse relaxation time (T_2), and the steady-state $\{^1H\}$-^{15}N NOE for each residue in the protein, information about the type and time scale of motion experienced by each residue can be obtained. Many motions on the ps-ns time scales with small excursion angles are regarded as librations or hindered motions within an energetic well and can be associated with retention of entropy in the folded state and with the protein's management of thermal energy in general. Slower motions are considered as activated processes in which the system must cross some energetic maximum and are very important because they are close to the time scales on which docking, protein folding, allosteric transitions, and catalysis take place.

Flexibility of DNA-contacting regions and its modulation upon DNA complex formation have now been reported for several DNA-binding proteins.[78−83] Often, the DBD in the free state is characterized by fast or slower time scale motions that quench upon binding to the specific DNA target. Thorough dynamic analysis of the binding of a certain protein to variant specific sequences would potentially hint at the role of dynamics in protein−DNA recognition; however, this area still remains unexplored. Very recently, the whole protein−DNA reaction process of the *lac* DBD has been dynamically characterized (Figure 11).[70] The data revealed for the first time how the various time-scale motions adjust along the protein−DNA reaction coordinate, from the free state to the specific complex via the nonspecific complex formation, and how the changes may modulate the thermodynamics of the system. More specifically, the heteronuclear NOEs for the α-helical residues were found to be >0.7 in all three forms, indicating that motions on the fast time scale (ps-ns) of structured regions are restricted and do not vary along the pathway (Figure 14A). In contrast, the loop residues (28−31) exhibited an interesting variation. In the free form the loop showed enhanced mobility on the ps-ns time scale, which becomes gradually more rigid upon binding to nonspecific and finally to specific DNA.[70,84] The conformation of this loop is important for recognition and binding stabilization; it interacts nonspecifically with DNA through hydrogen-bond and Coulombic interactions and recognizes the cognate sequence through specific hydrophobic contacts.

Slower motions on the μs-ms time scale exhibited a very interesting variation. Motions on this time scale are identified by the exchange term R_{ex}, which indicates exchange between conformations that sense different chemical environments.[83] Many residues in the free state show prominent μs-ms time-scale motions (Figure 14B). Binding to the nonspecific sequence increases significantly the R_{ex} term of most

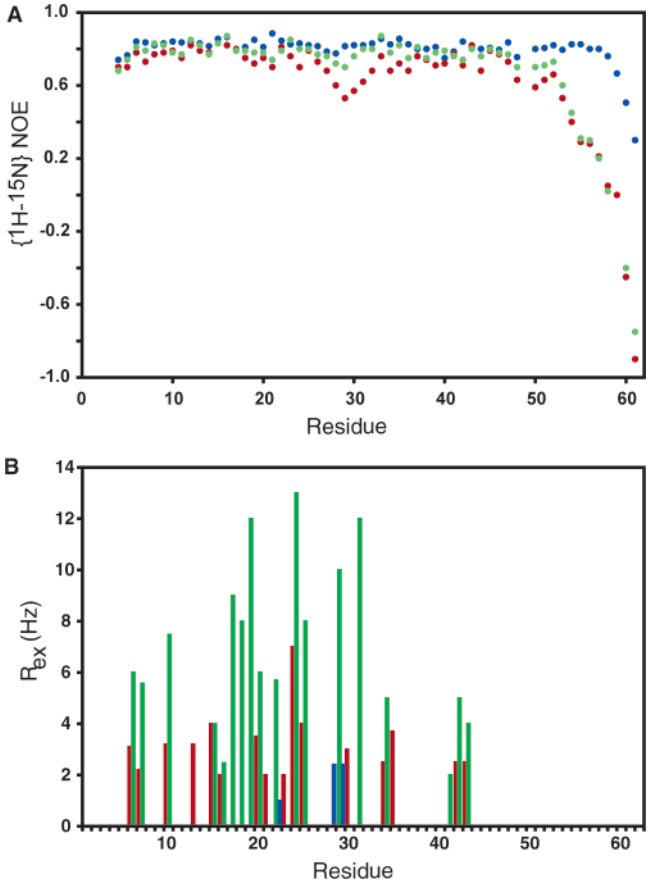


Figure 14. Dynamic analysis of protein−DNA recognition. In parts A and B values for the free state and nonspecific and specific complexes are shown in red, green, and blue, respectively. Results of (A) $\{^{15}N-^1H\}$ NOE and (B) exchange values R_{ex} indicating motions in the μs-ms time regime are plotted as a function of residue number of the *lac* HP62.

residues located in the protein−DNA interface, which show concerted motions on the μs-ms time scale. Of particular interest is the potential correlation between the dynamic features of the protein and its DNA-binding properties, as most of the residues that exhibit enhanced mobility on the μs-ms time scale form the protein−DNA interface and are involved in DNA binding in both the nonspecific and the specific complex. The structure ensemble analysis of the nonspecific complex revealed that while almost all of the residues with enhanced R_{ex} interact with the DNA backbone, some of them adopt alternative conformations.[70] For example, Arg22 and to a lesser extent Tyr17 and Gln18 were seen to participate not only in direct hydrogen bonds with the DNA phosphate but also in water-mediated interactions with base pairs in the major groove. From inspection of space-filling models of the structures, it seems that exchange between these positions is not hindered sterically. Thus, the flexibility enhancement may primarily originate from the sampling of different base-pair environments in the nonspecific complex, which is mainly performed by the recognition helix residues. The high flexibility of the complex on the μs-ms time scale may be of significant advantage for rapid and efficient location of the specific site, since these motions indicate thermally assisted conforma-

tional transitions. Interestingly, the prominent conformational dynamics exhibited by these residues in the free state and the nonspecifically bound state were totally quenched upon binding to the cognate operator, consistent with the selection of a single conformer (Figures 11B and 14B). These results support experimentally the view about nonspecific versus specific binding, where the less selective complexes are expected to have a more rugged bottom to the binding energy funnel, with low barriers between conformers of the complex.[85] This ruggedness corresponds to the presence of several conformers with slightly different energies, and hence, there is some interconversion between these conformers at a given temperature. This manifests itself as slow time scale dynamics. Each of these conformers is capable of binding ligands that are slightly different in conformation and thus are more forgiving to the detailed structural features. In contrast, highly specific complexes are relatively rigid, with a steep funnel of conformations leading to the native structure, and accommodate only those ligands which have distinct structural features.

5. Residue-Specific Insight into Energy Propagation and the Interaction Pathway in Protein–DNA Recognition

5.1. Redistribution of the Native-State Ensemble along the Protein–DNA Pathway

The propagation of energy through three-dimensional structure represents the physical basis for allosterism in biological systems. The precise pathway of signal transmission remains, however, elusive. Especially with respect to protein−DNA interactions, very little is known about the association and dissociation reaction pathways. The protein−DNA recognition process is quite intricate and involves formation of a large number of specific and nonspecific intermolecular contacts. In regulatory systems, the DNA-binding affinity of the protein, for example, the *lac* repressor, is modulated by binding of effector molecules at distant sites. Recently, it was proposed that an important condition for the propagation of binding effects to distal regions is the presence of a significant fraction of residues with low structural stability in the uncomplexed binding site.[86] Low intrinsic stability may be a common feature of regulatory proteins, thus challenging traditional concepts of the control of protein activity.[8,87,88] DNA binding may also induce allosteric changes in a protein that are crucial in the signal transduction process.[60] How these conformational changes are transmitted is of major importance for understanding the regulatory, kinetic, and recognition properties of the proteins.

Experimental observations, particularly those obtained from hydrogen−deuterium (H−D) exchange experiments, indicate that under native conditions the predominant equilibrium is not between the native state and the unfolded state but between a large number of states generated by the occurrence of local unfolding reactions within the native structure. These local unfolding reactions involve only a few residues, occur independently of each other, and define, to a large extent, the native-state ensemble.[89] The immediate and perhaps most relevant consequence is that the Gibbs energy of stabilization of a protein is not uniformly distributed throughout its 3D structure. According to the modern concepts of protein-folding funnels and energy landscapes, protein function is not determined purely by the static structure but rather through a redistribution of already existing populations in response to external perturbations.[90] Recently, it was demonstrated that NMR-detected exchange of amide protons can be used to characterize, at residue level, the mechanism, kinetics, and thermodynamics of the *lac* headpiece interaction with DNA operators, providing an unprecedented view of the interaction mechanism of a protein−DNA complex.[91] Furthermore, the redistribution of the native-state ensemble along the complete protein−DNA recognition pathway was monitored.

Hydrogen exchange techniques, with their residue-level specificity, exquisite sensitivity, and adaptability to many solution conditions, are becoming essential to the study of protein stability, folding and dynamics.[92−94] The exchange takes place according to the following scheme[95]

$$\text{closed (NH)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{open (NH)} \xrightarrow[D_2O]{k_{int}} \text{exchanged (ND)}$$

According to the model, exchange can take place only from the open conformation with an intrinsic rate constant k_{int}, which depends on sequence, pH, and temperature and can be easily calculated.[96] Open and closed conformations interconvert with the rate constants k_{op} and k_{cl}, respectively. There are two mechanisms by which exchange can take place: EX2 and EX1. In the first case the rate constant for reprotection, k_{cl}, is much greater than k_{int} and the observed hydrogen exchange reflects the equilibrium constant between the closed and the open states [$k_{obs} = (k_{op}/k_{cl})k_{int}$]. The ratio k_{cl}/k_{op} is referred to as the amide protection factor and can be used to estimate the free energy for the dominant opening reaction $\Delta G_{op} = -RT \ln(k_{obs}/k_{int})$. At the other extreme, when k_{int} is much greater than the rate constant for reprotection (typically above pH 9), the mechanism becomes EX1, and the observed rate constant becomes independent of k_{int} and simply equals the rate constant for the formation of the unprotected state, k_{op}. Thus, exchange rates measured under both conditions can be used to extract both the thermodynamics and kinetics of the opening event. Recently, it has been shown that this simple model for NH exchange appears to be a robust framework for obtaining quantitative information about molecular motions in native proteins.[97]

The *lac* system was used as a model to investigate with residue-level specificity the energy propagation upon DNA binding, as well as the intricate way a protein−DNA complex forms and dissociates. The hydrogen−deuterium exchange rates of each individual backbone amide proton of the dimeric *lac* headpiece in the free state and complexed to specific and nonspecific DNA operators were measured by

NMR spectroscopy. Thirty two protons per monomeric unit are protected in the free state, and an additional 5 and 12 are protected in the nonspecific and specific operator-bound states, respectively, due to participation in intermolecular hydrogen bonds, burial in the binding interface, or hinge helix formation. Protection factors for the *lac* DBD in the free state (Figure 15) were unexpectedly low (average

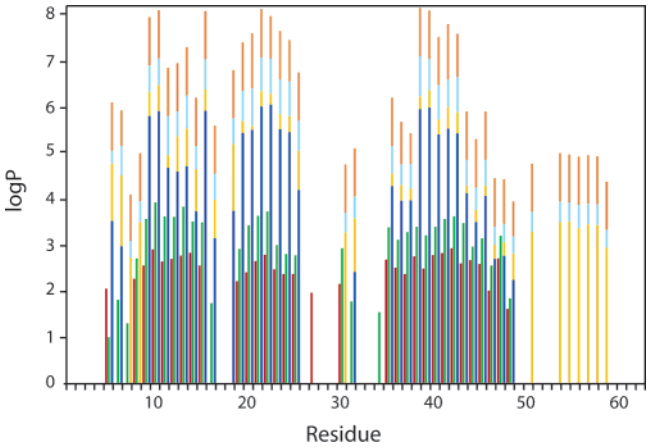


Figure 15. Protection factors (*P*) of the dimeric *lac* HP62 plotted as a function of residue number. Values for the free state, the nonspecific complex, the SymR complex, natural operators *O3* and *O1*, and the SymL complex are shown in red, green, blue, yellow, cyan, and orange, respectively.

value below 10^2) for a folded protein, suggesting a high population of open microstates (conformations from which exchange can take place). The low stability of the molecule in the uncomplexed state[98] was further corroborated by thermal denaturation experiments, which showed significant unfolding of the α-helices already at 42 °C. Interestingly, binding to the nonspecific DNA fragment results in a small increase in the protection factors; thus, no significant redistribution of the protein ensemble takes place. However, upon binding to the SymL operator (Figure 14) there is a dramatic increase in protection as a result of consolidation of the complex. Thus, binding to the cognate operator reduces dramatically the population of open conformational states of the *lac* DBD, narrowing the population distribution and contributing significantly to the observed negative heat capacity (see below). Interestingly, it was demonstrated that the pattern and magnitude of protection factors of *lac* DBD binding to DNA are modulated by the degree of specificity of the operator sequence (Figure 15). SymL is the sequence that binds more strongly to the repressor, followed by the natural operators *O1*, *O2*, and *O3*. As Figure 15 depicts, the extent of population narrowing follows the specificity trend. The natural *lac* operator *O1* (Figure 2) is asymmetric, and the two binding sites differ significantly in their affinity for the *lac* repressor when considered separately.[99] Quite interestingly, measured protection factors for the dimeric *lac* headpiece complexed to its cognate operator are the same for the left and the right protein subunit (Figure 14B). The significant cooperative effect of hinge helices stabilization on the headpiece core is demonstrated in the case of the SymR operator, where the hinge helices exist at equilibrium between α-helical and random coil conformations (Figure 4C). This complex exhibits the lowest protection factors among the specific sequences. Therefore, the *lac* repressor employs the same binding site to interact with both specific and nonspecific DNA sequences, with the equilibrium population of the protein conformational substates being redistributed and shifted depending upon the DNA sequence. The data suggested a hierarchical model where the biological system becomes more ordered going from the nonspecific to the most specific complex formation.

Even more impressive was the observation that protection is not confined to the binding interface but is spread throughout the whole molecule, even for nonspecific binding. This is clearly a manifestation of extended energetic coupling between different regions of the protein. The locally resolved free energy changes in protein upon DNA binding are depicted in Figure 15. How can the dramatic DNA-induced redistribution of the native-state ensemble be explained? Because the binding domain is not intrinsically stable, DNA binding selects those states in which the core is stably formed. This process results in a large redistribution of the native-state ensemble in the presence of the DNA. Because the DNA-induced redistribution of the native-state ensemble affects not only those regions in direct contact with the DNA but also regions linked by cooperative interactions, the potential exists for the occurrence of effects at distal sites from the binding site.

5.2. Residue-Specific View of the Dissociation and Association Pathways of the Protein–DNA Complex

The lac repressor−operator system was used as a model for developing a methodology with the exciting prospect of following the dissociation of the protein−DNA complex with residue-level specificity. To this end, hydrogen exchange rates of the dimeric *lac* DBD bound to SymL operator were measured as a function of pH. For local exchange, EX2 kinetics were observed as expected, since the native-state fluctuations are rapid and there is a fast closing step (Figure 16A). However, most of the residues switched to the EX1 mechanism at elevated pH values. The possibility of following the kinetics of structural fluctuations required to break the hydrogen bond and/or expose individual amides to solvent is of great importance, since it may reveal concerted phenomena that are not possible to follow with conventional experiments. The results are mapped on the three-dimensional structure of the complex in Figure 16B. Opening rates segregate in four distinct groups. The first group to open is the hinge helix, with an exchange rate of 0.20 h^{-1}, followed by N50 and the C-terminal residues of the third helix (0.11 h^{-1}). Unfolding of this part of the protein is propagated to the rest of the molecule by progressive destabilization and collapse of protein−protein and protein−DNA contacts, resulting in the complex being dissociated with a rate of 0.02 h^{-1}. The apparent dissociation constant of the complex, measured by biochemical methods, is ∼0.05 h^{-1},[35] indicating that the slowest k_{op} rate corresponds to the

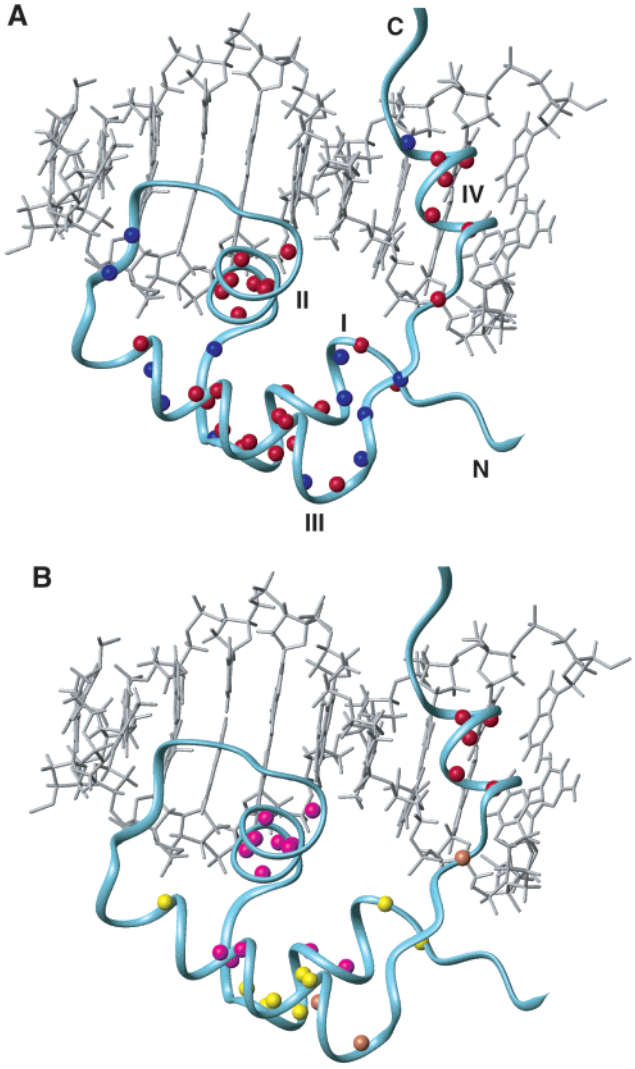


Figure 16. Summary of the kinetic data extracted from hydrogen exchange data. For clarity reasons, only the left site of the complex is depicted. Helices II and IV refer to the recognition and hinge helices, respectively. (A) All backbone amide protons that are protected in the DNA-bound state. NHs that exchange only under the EX2 mechanism are blue, whereas those that switch to the EX1 mechanism at high pH values are red. (B) Mapping of the opening rates, k_{op}, of the dimeric *lac* HP62 bound to the SymL operator. The color code used to display k_{op} is as follows: $k_{op} = 0.20\ h^{-1}$, red; $k_{op} = 0.11\ h^{-1}$, coral; $k_{op} = 0.04\ h^{-1}$, yellow; $k_{op} = 0.02\ h^{-1}$, violet. Reprinted with permission from ref 89. Copyright 2002 Nature Publishing Group.

macroscopic kinetics of dissociation. The group of backbone amide hydrogens slowest to exchange tend to cluster in mutually packed elements of secondary structure, which suggests the existence of a partially folded submolecular domain. Its structure is stabilized by the specific contacts to DNA from the recognition helix, which remains intact, and by strong hydrophobic interactions with residues from the other helices (Figure 16B). On the basis of the variation of the lifetime of protected states, a model of progressive unfolding and dissociation of the protein−DNA complex was proposed and is illustrated in Figure 17. Although identical k_{op} rates do not necessarily imply that cooperative unfolding takes place, binding of the *lac* repressor to DNA is accompanied by cooperative phenomena. For example, formation of the hinge helix, which has been shown to require both protein−protein and protein−DNA contacts, is stable only when the hinge helix of the other subunit is folded. The recognition helix should be the last one to dissociate from DNA, because the side chains of its residues form most of the contacts to the operator. The dissociation pathway of the *lac* repressor from its wild-type operator was found to be identical to that of the symmetric operator, yet the complex responds to hinge helix unfolding twice as fast as the SymL complex. It is remarkable that measurement of individual opening rates can reveal with such detail the pathway of the progressive unfolding of subglobal structural units, providing a clear picture of how the protein dissociates from DNA.

The H−D approach followed in that work revealed further important information about the association pathway of the protein−DNA complex as well. The key to accomplishing this was to measure the closing rates, namely the rates by which amide protons become reprotected. As residues that are protected only in the DNA-bound state were considered, the potential existed of monitoring the process of reprotection of individual structural units. The analysis showed that Gln18, a crucial residue of the recognition helix, becomes solvent protected ~2 times faster than Leu6, with a rate of $\sim 40\ s^{-1}$ (this is consistent with an apparent association rate constant of $\sim 10^5\ M^{-1}\ s^{-1}$). Based also on the submolecular core that is slowest to exchange, the results suggest that the recognition helix is the first to fit to DNA and this orientation is locked by subsequent hydrogen bond formation of the Leu6 backbone to DNA. It is interesting to note that this contact is highly conserved among helix-turn-helix (HTH) proteins, irrespective of the nature of the residue.[19] Hinge helix formation is slower than Qln18 protection by a factor of ~10, and the last H-bond to form is that of Asn50 with the DNA phosphate. This observation suggests that protein−DNA interactions in the minor groove are difficult to establish and can take place only after the HTH core of the headpiece has been properly oriented with respect to DNA. Thus, the DNA sequence is first read out in the major groove by the recognition helix, followed by discrimination of the minor groove by the hinge region and its subsequent folding to an α-helix. This view is enhanced by the recent structural data of the nonspecific complex, which showed that while the headpiece has positioned its recognition helix over the major groove of the operator and Leu6 is hydrogen bonded to the DNA backbone, the hinge region is unstructured and folds up to an α-helix only when the specific sequence is encountered.

5.3. Implications for the Allosteric Mechanism and DNA Recognition

The significant instability of the DNA-binding domain in the uncomplexed state appears to have severe implications for the allosteric mechanism whereby the intact *lac* repressor exerts control over gene expression. If only "high-stability" residues constituted the binding site, then all the states in the

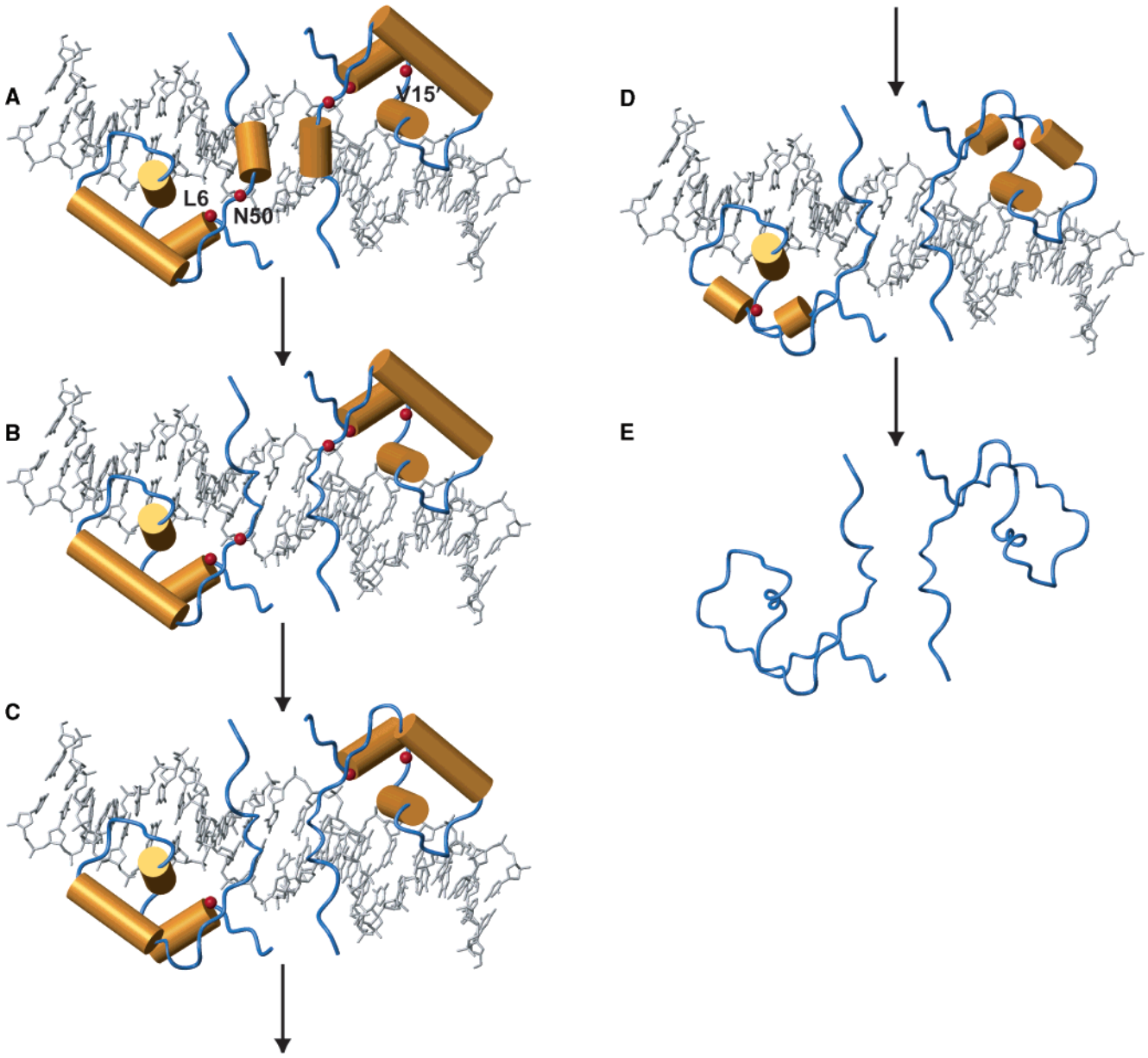


Figure 17. Model of the dissociation pathway of the *lac* repressor from the ideal SymL operator. The progressive unfolding of individual structural units is based on the opening rates measured by hydrogen−deuterium exchange for the individual sites. Leu6 and Asn50 hydrogen bond to DNA phosphate, whereas the backbone amide of Val15 hydrogen bonds to the carbonyl oxygen of Ala10 and is a key residue in defining the proper orientation of the HTH domain. (A) All rarely opened residues are protected. The first structural unit to unfold is that of the hinge helices (B) followed by disruption of the N50 to DNA hydrogen bond and opening of the C-terminal residues of the third helix (C). Further unfolding takes place in part D, where the hydrogen bond from L6 to DNA is disrupted and the only helix that remains intact is the recognition helix. In part D, the slowest to exchange core remains that is stabilized by specific contacts to DNA from the recognition helix and strong hydrophobic interactions with residues from the other helices. V15 and the residues located in the middle of the first and third helices do not contact DNA but are apparently very important because they are involved in critical packing interactions with the recognition helix that stabilize this submolecular domain. Their disruption results in protein−DNA dissociation (E). Reprinted with permission from ref 89. Copyright 2002 Nature Publishing Group.

ensemble would be binding-competent, and DNA binding would induce only an energy shift without any redistribution of the states in the ensemble.[86] In that case, unfolding of the hinge helices would affect only binding to the minor groove. The large difference in mobility between the free and DNA-bound *lac* DBDs can be used advantageously by the biological system to switch between functional states. Therefore, alteration of dynamics between free and bound states coupled to destabilization of crucial structural subunits may provide a level of control that allows rapid and accurate response of the biological system to changes in the cell environment.

As has been demonstrated, two important phenomena accompany recognition of sequence-specific operators by the *lac* repressor: hinge helix formation and a profound reduction in the manifold of the protein native-state ensemble. The slow time-scale flexibility of the DBD of the *lac* repressor in the presence of noncognate operator sites allows it to slide along the DNA facilitating target location. Once the specific site has been recognized in the major

groove, the hinge helices form and recognize the minor groove. The recognition is followed by selection of only the binding-competent states of the ensemble, which manifests itself as a negative contribution to the heat capacity of the system (see below). The combined structural and dynamic results provide a unique view of the order of events that result in protein−DNA association and dissociation, which may enrich our understanding of how proteins proceed toward selecting their target sites from the sea of nonspecific sequences present in the environment. Moreover, they present a compelling example of a large redistribution in the ensemble of conformational states of a protein when it binds specifically to DNA.

6. Relating Structure and Dynamics with Thermodynamics

6.1. Variation of ΔC_p in Protein–DNA Interactions

To provide a proper molecular foundation for the elucidation of the protein−DNA recognition features, an in-depth insight into the energetics of the process is needed. The *lac* repressor−operator system may again serve as a model system for which a correlation can be made between the structure, dynamics, and thermodynamics of the biological process. A large body of thermodynamic data exists for the binding of both the intact *lac* repressor and its isolated headpiece to various operator sequences.[100] 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 the nonspecific binding.[69] Thus, many efforts have been focused on understanding heat capacity, a fundamental thermodynamic parameter, and correlating its magnitude and sign of change with protein−DNA binding parameters. ΔC_p for the *lac* repressor binding to specific operators is $\approx -1100-1300$ cal mol^{-1} K^{-1}, whereas for nonspecific binding it is only ≈ -200 cal mol^{-1} K^{-1}.[100] The most productive link between thermodynamic and structural data has proved to be the relationship between ΔC_p and the change in buried surface area on going from one equilibrium state to another. The variation in ΔC_p of specific versus nonspecific binding has been generally accounted for by a difference in the solvent-accessible surface area (ASA) removed from bulk water upon complex formation. In light of the recently available structural and dynamic data, a molecular interpretation of ΔC_p can be attempted. Binding to the cognate operator *O1* results in 4900 Å^2 (2500 Å^2 nonpolar; 2400 Å^2 polar) of buried surface area compared to 3500 Å^2 (2100 Å^2 nonpolar; 1400 Å^2 polar) for the nonspecific one. The contribution of surface area burial to ΔC_p amounts to $\sim$465 cal mol^{-1} K^{-1} for the specific complex and to a maximum of $\sim$470 cal mol^{-1} K^{-1} for the nonspecific one (using the algorithm $\Delta C_p = -(0.34 \pm 0.04)\Delta A_{np} + (0.14 \pm 0.04)\Delta A_p$ cal mol^{-1} K^{-1}).[69] Although formation of the specific complex results in a 1400 Å^2 excess of the buried surface, its contribution to ΔC_p is the same as that of the nonspecific one. This is due to the much higher amount of polar surface being excluded from the solvent in the specific complex, which contributes to a positive ΔC_p. Thus, the portion of ΔC_p left unexplained by surface area considerations for specific binding is more than 600 cal mol^{-1} K^{-1}. The failure to account for the total observed ΔC_p solely on buried surface arguments has been observed very often.

There are, however, additional sources of ΔC_p that have proved very difficult to integrate in a model due to scarce data on nonspecific binding.[101] Changes in flexibility upon binding modulate conformational entropy, which contributes significantly to ΔC_p. Both ps-ns and μs-ms time scale fluctuations, which can be probed by relaxation experiments, affect ΔC_p. Furthermore, restriction of soft vibrational modes characterized by force constants weak enough to be affected by ligand binding has been suggested to account for up to 20% of the total ΔC_p.[102] The dynamic data demonstrate that the protein−DNA interface in the nonspecific complex is highly flexible on all probed time scales, whereas in the specific complex it is tied down by strong contacts and optimal juxtaposition (Figures 11 and 14). Moreover, the redistribution of the native-state ensemble induced by ligand binding appears to be a significant source of ΔC_p. As has been earlier suggested, reduction of the population of microstates will give rise to increased negative ΔC_p.[103,104] Our H−D exchange data reveal that *lac* binding to the nonspecific operator has a small effect on the ensemble, whereas binding to the specific operator results in dramatic narrowing of the population distribution throughout the whole molecule. Additionally, the extent of narrowing is modulated by the degree of specificity of the DNA sequence and reduces in the order SymL > *O1* > *O3* (Figure 15). ΔC_p of both the intact repressor[69,100] and the isolated dimeric HP62[70] binding also significantly decreases in the same order. As both burial of surface area[22,41] and relaxation data (Figure 11)[70] are identical for these complexes, it becomes apparent that the distribution of the native-state ensemble significantly affects ΔC_p.

6.2. Salt Dependence of Specific versus Nonspecific Binding

The available data further provided the proper structural basis for explaining another long-standing issue in protein−DNA interactions: the large and characteristically different salt dependences of specific versus nonspecific binding. Comparison of the high-resolution structures of the nonspecific and specific complexes reveals that significant charge redistribution takes place, which results in stronger electrostatic interactions in the nonspecific complex. Formation of the nonspecific complex involves the release of $\sim$11 monovalent ions while formation of the specific complex releases the thermodynamic equivalent of only $\sim$6 ions from the vicinity of the DNA.[105] His29, Lys33, and Lys59 are the only charged residues whose side chains are close to DNA phosphates in the specific complex (Figure 7), thus giving rise to three ion pairs per monomer (six in total). Analysis of the nonspecific structure suggests that (additionally to His29 and Lys33) Arg22, Arg35, and Lys37 interact with the DNA phosphate, whereas

Lys59 moves away due to the unfolding of the hinge helices, accounting for the two additional charge−charge interactions that are expected on the basis of the thermodynamic findings.

The combined set of structural and dynamic data on the DNA-binding domain of the *lac* repressor summarized in the present review provides a very detailed insight into how a protein progresses from its free state toward selecting its cognate DNA site from the sea of nonspecific sequences present in the environment and into the mechanisms that the biological system makes use of in order to accomplish its task.

7. References

(1) Jen-Jacobson, L. *Biopolymers* **1997**, *44*, 153.
(2) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804.
(3) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. *Nucleic Acids Res.* **2001**, *29*, 2860.
(4) Pabo, C. O.; Nekludova, L. *J. Mol. Biol.* **2000**, *301*, 597.
(5) Garvie, C. W.; Wolberger, C. *Mol. Cell* **2001**, *8*, 937.
(6) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. *Nature* **1988**, *335*, 321.
(7) Billeter, M.; Güntert, P.; Luginbuhl, P.; Wüthrich, K. *Cell* **1996**, *85*, 1057.
(8) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321.
(9) von Hippel, P. H. *Science* **1994**, *263*, 769.
(10) Halford, S. E.; Marko, J. F. *Nucleic Acids Res.* **2004**, *32*, 3040.
(11) Bell, C. E.; Lewis, M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 19.
(12) Pace, H. C.; Kercher, M. A.; Lu, P.; Markiewicz, P.; Miller, J. H.; Chang, G.; Lewis, M. *Trends Biochem. Sci.* **1997**, *22*, 334.
(13) Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P. *Science* **1996**, *271*, 1247.
(14) Bell, C. E.; Lewis, M. *Nature Struct. Biol.* **2000**, *7*, 209.
(15) Bell, C. E.; Lewis, M. *J. Mol. Biol.* **2001**, *312*, 921.
(16) Ogata, R. T.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5851.
(17) Boelens, R.; Scheek, R. M.; van Boom, J. H.; Kaptein, R. *J. Mol. Biol.* **1987**, *193*, 213.
(18) Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland: New York, 1999.
(19) Wintjens, R.; Rooman, M. *J. Mol. Biol.* **1996**, *262*, 294.
(20) Schumacher, M. A.; Choi, K. Y.; Zalkin, H.; Brennan, R. G. *Science* **1994**, *266*, 763.
(21) Spronk, C. A.; Slijper, M.; van Boom, J. H.; Kaptein, R.; Boelens, R. *Nature Struct. Biol.* **1996**, *3*, 916.
(22) Spronk, C. A.; Bonvin, A. M.; Radha, P. K.; Melacini, G.; Boelens, R.; Kaptein, R. *Structure* **1999**, *7*, 1483.
(23) Nagadoi, A.; Morikawa, S.; Nakamura, H.; Enari, M.; Kobayashi, K.; Yamamoto, H.; Sampei, G.; Mizobuchi, K.; Schumacher, M. A.; Brennan, R. G.; et al. *Structure* **1995**, *3*, 1217.
(24) Werner, M. H.; Huth, J. R.; Gronenborn, A. M.; Clore, G. M. *Cell* **1995**, *81*, 705.
(25) Murphy, F. V. t.; Sweet, R. M.; Churchill, M. E. *EMBO J.* **1999**, *18*, 6610.
(26) Heldwein, E. E.; Brennan, R. G. *Nature* **2001**, *409*, 378.
(27) Markiewicz, P.; Kleina, L. G.; Cruz, C.; Ehret, S.; Miller, J. H. *J. Mol. Biol.* **1994**, *240*, 421.
(28) Sadler, J. R.; Sasmor, H.; Betz, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6785.
(29) Zinkel, S. S.; Crothers, D. M. *Nature* **1987**, *328*, 178.
(30) Perros, M.; Steitz, T. A. *Science* **1996**, *274*, 1929.
(31) Fried, M. G.; Hudson, J. M. *Science* **1996**, *274*, 1930.
(32) Spronk, C. A.; Folkers, G. E.; Noordman, A. M.; Wechselberger, R.; van den Brink, N.; Boelens, R.; Kaptein, R. *EMBO J.* **1999**, *18*, 6472.
(33) Suckow, J.; Markiewicz, P.; Kleina, L. G.; Miller, J.; Kisters-Woike, B.; Muller-Hill, B. *J. Mol. Biol.* **1996**, *261*, 509.
(34) Chakerian, A. E.; Matthews, K. S. *J. Biol. Chem.* **1991**, *266*, 22206.
(35) Kalodimos, C. G.; Folkers, G. E.; Boelens, R.; Kaptein, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6039.
(36) Falcon, C. M.; Swint-Kruse, L.; Matthews, K. S. *J. Biol. Chem.* **1997**, *272*, 26818.
(37) Sauer, R. T.; Hehir, K.; Stearman, R. S.; Weiss, M. A.; Jeitler-Nilsson, A.; Suchanek, E. G.; Pabo, C. O. *Biochemistry* **1986**, *25*, 5992.
(38) Simoncsits, A.; Chen, J.; Percipalle, P.; Wang, S.; Toro, I.; Pongor, S. *J. Mol. Biol.* **1997**, *267*, 118.
(39) Robinson, C. R.; Sauer, R. T. *Biochemistry* **1996**, *35*, 109.
(40) Jamin, N.; Toma, F. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 83.
(41) Kalodimos, C. G.; Bonvin, A. M.; Salinas, R. K.; Wechselberger, R.; Boelens, R.; Kaptein, R. *EMBO J.* **2002**, *21*, 2866.
(42) Falcon, C. M.; Matthews, K. S. *Biochemistry* **2000**, *39*, 11074.
(43) Ogata, R. T.; Gilbert, W. *J. Mol. Biol.* **1979**, *132*, 709.
(44) Betz, J. L.; Sasmor, H. M.; Buck, F.; Insley, M. Y.; Caruthers, M. H. *Gene* **1986**, *50*, 123.
(45) Rastinejad, F.; Artz, P.; Lu, P. *J. Mol. Biol.* **1993**, *233*, 389.
(46) Horton, N.; Lewis, M.; Lu, P. *J. Mol. Biol.* **1997**, *265*, 1.
(47) Falcon, C. M.; Matthews, K. S. *Biochemistry* **2001**, *40*, 15650.
(48) Sartorius, J.; Lehming, N.; Kisters, B.; von Wilcken-Bergmann, B.; Muller-Hill, B. *EMBO J.* **1989**, *8*, 1265.
(49) Luisi, B. F.; Xu, W. X.; Otwinowski, Z.; Freedman, L. P.; Yamamoto, K. R.; Sigler, P. B. *Nature* **1991**, *352*, 497.
(50) Gewirth, D. T.; Sigler, P. B. *Nature Struct. Biol.* **1995**, *2*, 386.
(51) Rodgers, D. W.; Harrison, S. C. *Structure* **1993**, *1*, 227.
(52) Chen, Y. Q.; Sengchanthalangsy, L. L.; Hackett, A.; Ghosh, G. *Structure* **2000**, *8*, 419.
(53) Ogata, R.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 4973.
(54) Goeddel, D. V.; Yansura, D. G.; Caruthers, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3378.
(55) Betz, J. L.; Sasmor, H. M.; Buck, F.; Insley, M. Y.; Caruthers, M. H. *Gene* **1986**, *50*, 123.
(56) Lehming, N.; Sartorius, J.; Niemoller, M.; Genenger, G.; v Wilcken-Bergmann, B.; Muller-Hill, B. *EMBO J* **1987**, *6*, 3145.
(57) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163.
(58) Drew, H. R.; Travers, A. A. *Nucleic Acids Res.* **1985**, *13*, 4445.
(59) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. *Structure* **2000**, *8*, 1015.
(60) Lefstin, J. A.; Yamamoto, K. R. *Nature* **1998**, *392*, 885.
(61) Koudelka, G. B. *Nucleic Acids Res.* **1998**, *26*, 669.
(62) Shimamoto, N. *J. Biol. Chem.* **1999**, *274*, 15293.
(63) von Hippel, P. H.; Berg, O. G. *J. Biol. Chem.* **1989**, *264*, 675.
(64) Riggs, A. D.; Bourgeois, S.; Cohn, M. *J. Mol. Biol.* **1970**, *53*, 401.
(65) Kabata, H.; Kurosawa, O.; Arai, I.; Washizu, M.; Margarson, S. A.; Glass, R. E.; Shimamoto, N. *Science* **1993**, *262*, 1561.
(66) Misteli, T. *Science* **2001**, *291*, 843.
(67) Gowers, D. M.; Halford, S. E. *EMBO J.* **2003**, *22*, 1410.
(68) Jeltsch, A.; Wenz, C.; Stahl, F.; Pingoud, A. *EMBO J.* **1996**, *15*, 5104.
(69) Spolar, R. S.; Record, M. T., Jr. *Science* **1994**, *263*, 777.
(70) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M.; Guennuegues, M.; Boelens, R.; Kaptein, R. *Science* **2004**, *305*, 386.
(71) Mossing, M. C.; Record, M. T., Jr. *J. Mol. Biol.* **1985**, *186*, 295.
(72) Winkler, F. K.; Banner, D. W.; Oefner, C.; Tsernoglou, D.; Brown, R. S.; Heathman, S. P.; Bryan, R. K.; Martin, P. D.; Petratos, K.; Wilson, K. S. *EMBO J.* **1993**, *12*, 1781.
(73) Viadiu, H.; Aggarwal, A. K. *Mol. Cell* **2000**, *5*, 889.
(74) Forman-Kay, J. D. *Nature Struct. Biol.* **1999**, *6*, 1086.
(75) Cavanagh, J.; Akke, M. *Nature Struct. Biol.* **2000**, *7*, 11.
(76) Ishima, R.; Torchia, D. A. *Nature Struct. Biol.* **2000**, *7*, 740.
(77) Kay, L. E. *Nature Struct. Biol.* **1998**, *5 Suppl*, 513.
(78) van Tilborg, P. J.; Czisch, M.; Mulder, F. A.; Folkers, G. E.; Bonvin, A. M.; Nair, M.; Boelens, R.; Kaptein, R. *Biochemistry* **2000**, *39*, 8747.
(79) Mulder, F. A.; van Tilborg, P. J.; Kaptein, R.; Boelens, R. *J. Biomol. NMR* **1999**, *13*, 275.
(80) Vaughn, J. L.; Feher, V. A.; Bracken, C.; Cavanagh, J. *J. Mol. Biol.* **2001**, *305*, 429.
(81) Wikstrom, A.; Berglund, H.; Hambraeus, C.; van den Berg, S.; Hard, T. *J. Mol. Biol.* **1999**, *289*, 963.
(82) Hard, T. *Q. Rev. Biophys.* **1999**, *32*, 57.
(83) Akke, M. *Curr. Opin. Struct. Biol.* **2002**, *12*, 642.
(84) Slijper, M.; Boelens, R.; Davis, A. L.; Konings, R. N.; van der Marel, G. A.; van Boom, J. H.; Kaptein, R. *Biochemistry* **1997**, *36*, 249.
(85) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng.* **1999**, *12*, 713.
(86) Freire, E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10118.
(87) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graphics Modell.* **2001**, *19*, 26.
(88) Gunasekaran, K.; Tsai, C. J.; Kumar, S.; Zanuy, D.; Nussinov, R. *Trends Biochem. Sci.* **2003**, *28*, 81.
(89) Englander, S. W. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 213.

(90) Tsai, C. J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181.
(91) Kalodimos, C. G.; Boelens, R.; Kaptein, R. *Nature Struct. Biol.* **2002**, *9*, 193.
(92) Englander, S. W.; Krishna, M. M. *Nature Struct. Biol.* **2001**, *8*, 741.
(93) Chamberlain, A. K.; Handel, T. M.; Marqusee, S. *Nature Struct. Biol.* **1996**, *3*, 782.
(94) Parker, M. J.; Dempsey, C. E.; Hosszu, L. L.; Waltho, J. P.; Clarke, A. R. *Nature Struct. Biol.* **1998**, *5*, 194.
(95) Hvidt, A.; Nielsen, S. O. *Adv. Protein Chem.* **1966**, *21*, 287.
(96) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17*, 75.
(97) Sivaraman, T.; Arrington, C. B.; Robertson, A. D. *Nature Struct. Biol.* **2001**, *8*, 331.
(98) Felitsky, D. J.; Record, M. T., Jr. *Biochemistry* **2003**, *42*, 2202.
(99) Sasmor, H. M.; Betz, J. L. *Gene* **1990**, *89*, 1.
(100) Frank, D. E.; Saecker, R. M.; Bond, J. P.; Capp, M. W.; Tsodikov, O. V.; Melcher, S. E.; Levandoski, M. M.; Record, M. T., Jr. *J. Mol. Biol.* **1997**, *267*, 1186.
(101) Sturtevant, J. M. *Proc Natl. Acad. Sci. U.S.A.* **1977**, *74*, 2236.
(102) Ladbury, J. E.; Wright, J. G.; Sturtevant, J. M.; Sigler, P. B. *J. Mol. Biol.* **1994**, *238*, 669.
(103) Eftink, M. R.; Anusiem, A. C.; Biltonen, R. L. *Biochemistry* **1983**, *22*, 3884.
(104) Wang, C.; Pawley, N. H.; Nicholson, L. K. *J. Mol. Biol.* **2001**, *313*, 873.
(105) Record, M. T., Jr.; deHaseth, P. L.; Lohman, T. M. *Biochemistry* **1977**, *16*, 4791.
(106) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graphics* **1996**, *14*, 51.

CR0304065