Ligand–induced Conformational Changes and Conformational Dynamics in the Solution Structure of the Lactose Repressor Protein

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · MARCH 2008

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2007.11.067 · Source: PubMed

7 AUTHORS, INCLUDING:

Marc B Taraban
University of Maryland, Baltimore
76 PUBLICATIONS **636** CITATIONS
SEE PROFILE

Hongli Zhan
University of Alabama at Birmingham
19 PUBLICATIONS **280** CITATIONS
SEE PROFILE

David Langley
VCCRI and Garvan
47 PUBLICATIONS **742** CITATIONS
SEE PROFILE

Kathleen S Matthews
Rice University
168 PUBLICATIONS **3,352** CITATIONS
SEE PROFILE

Published in final edited form as:
J Mol Biol. 2008 February 15; 376(2): 466–481.

Ligand-Induced Conformational Changes and Conformational Dynamics in the Solution Structure of the Lactose Repressor Protein

Marc Taraban[1,†], **Hongli Zhan**[2,3], **Andrew E. Whitten**[4,5], **David B. Langley**[5], **Kathleen S. Matthews**[2], **Liskin Swint-Kruse**[6,*], and **Jill Trewhella**[1,5,*]

1 *Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA*

2 *Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77005, USA*

3 *Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, Kansas City, KS 66160, USA*

4 *Bragg Institute, Australian Nuclear Science and Technology Organization, NSW, Menai 2234 Australia*

5 *School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia*

6 *Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, Kansas City, KS 66160, USA*

SUMMARY

We present here the results of a series of small-angle X-ray scattering studies aimed at understanding the role of conformational changes and structural flexibility in DNA binding and allosteric signaling in a bacterial transcription regulator, Lactose repressor protein (LacI). Experiments were designed to detect possible conformational changes that occur when LacI binds either DNA or the inducer IPTG, or both. Our studies included the native LacI dimer of homodimers and a dimeric variant (R3), enabling us to probe conformational changes within the homodimers and distinguish them from those involving changes in the homodimer-homodimer relationships. The scattering data indicate that removal of operator DNA (*o*DNA) from R3 results in an unfolding and extension of the hinge-helix that connects the LacI regulatory and DNA-binding domains. In contrast, only very subtle conformational changes occur in the R3 dimer-*o*DNA complex upon IPTG binding, indicative of small adjustments in the orientations of domains and/or sub-domains within the structure. The binding of IPTG to native (tetrameric) LacI-*o*DNA complexes also appears to facilitate a modest change in the average homodimer-homodimer disposition. Notably, the crystal structure of the native LacI-*o*DNA complex differs significantly from the average solution conformation. The solution scattering data are best-fit by an ensemble of structures that includes (1) ~60% of the V-shaped dimer-of-homodimers observed in the crystal structure, and (2) ~40% of molecules with more "open" forms, such as those generated when the homodimers move with respect to each other about the tetramerization domain. In gene regulation, such a flexible LacI would be beneficial for the interaction of its two DNA binding domains, positioned at the tips of the V, with the required two of three LacI operators needed for full repression.

*Corresponding authors. E-mail addresses: lswint-kruse@kumc.edu; jtrewhella@usyd.edu.au.
†*Current address.* Institute of Chemical Kinetics and Combustion, Novosibirsk, 630090 Russia.

Keywords

LacI; allostery; small; angle X; ray scattering; solution structure

INTRODUCTION

Conformational flexibility and disordered structures play crucial roles in protein function, often controlling critical steps in recognition and binding. However, flexible and disordered structures are a challenge to characterize, because they often inhibit crystal formation. Even when crystals do form, the resolution of the diffraction data is limited and flexible or disordered regions may not be resolved in electron density maps. While NMR can probe dynamic structures, the size limitations mean that these studies frequently must use truncated forms of proteins. As discussed below, both flexibility and disorder have been postulated as functionally important for the 154,520 Da, tetrameric bacterial transcription regulator, Lactose repressor protein (LacI).

LacI binds DNA as a homodimer *via* a pair of N-terminal DNA-Binding Domains (DBDs), each of which is linked by eighteen amino acids to a regulatory domain (Figure 1A) [1; 2; 3; 4; 5]. Most homologues of LacI are homodimers that bind one operator DNA binding site (*o*DNA)[6]. A few – including LacI – have additional C-terminal tetramerization domains [7; 8; 9]. Tetramer formation allows binding to two *o*DNA sites and results in tighter repression (10, reviewed in [11]). Transcription repression is modulated when each regulatory domain binds an inducer molecule (two per dimer) which allosterically diminishes the protein's affinity for DNA [12; 13]. The natural inducer for LacI is allolactose (a structural isomer of lactose), and isopropyl β-D-1-thiogalactopyranoside (IPTG) is a gratuitous inducer [13; 14; 15]. Upon induction, nonspecific DNA (*n*DNA) effectively competes with *o*DNA for LacI binding [16; 17; 18].

Thus, the functional cycle of LacI comprises at least the following species: apo-LacI, *o*DNA bound LacI (LacI-*o*DNA), inducer-bound LacI, the ternary complex of LacI-*o*DNA-inducer, and LacI complexes with *n*DNA with and without bound inducer. Crystal structures are available for apo-LacI, LacI-*o*DNA, and IPTG-bound LacI [3; 4; 5; 19; 20], and NMR structures are available for the truncated DBD (*e.g.*[21; 22; 23; 24; 25]). However, no single structure is complete and all the crystal structures are low-to-moderate resolution, in part a result of motions that occur between different structural elements. These structures suggest three types of motion; between the homodimers, between the regulatory subdomains, and between the DBD and regulatory domain (Figure 1A).

The structure of the LacI tetramer bound to two operators shows the protein as a V-shaped dimer of homodimers (Figure 1A). Although all domains are present in this structure, the resolution (4.8 Å) is such that only the positions of Cα atoms are approximated [3]. Better resolution (2.6 Å) was obtained using a truncated, dimeric version of LacI, suggesting that the resolution of the tetramer structure is limited by motions between the homodimers [5]. The inducer binding site in the LacI regulatory domains sits in a central pocket between two subdomains, and the subdomains adopt different orientations in the DNA- and IPTG-bound structures [3; 5; 20; 26]. Finally, when bound to DNA, the LacI DBD and linkers are ordered and the middle portion of the linker folds into an α-helix (termed the "hinge helix") [5]. In the IPTG-bound state, no electron density is present for either the DBD or linkers [3]. Several additional lines of evidence – including NMR structures of the truncated DBD – suggest that the lack of electron density can be attributed to unfolding of the hinge helix and an accompanying increase in mobility [27; 28; 29; 30]. However, this unfolding has not been directly detected for full-length LacI in either the dimeric or tetrameric forms.

The observed regulatory subdomain reorientation and the putative unfolding of the linker hinge helix both have the potential to participate in the allosteric conformational change that alleviates repression. *In vivo,* the induced LacI complexes include both the ternary complex LacI-*o*DNA-IPTG complex and LacI-IPTG bound to *n*DNA. Nothing is known about how the conformation of the linkers, including the hinge helices, might change when the LacI-*o*DNA complex binds IPTG. One possibility is that the subdomain motions pull apart the hinge helices, causing them to unfold. Thus, several direct contacts to DNA would be abolished, resulting in diminished affinity. Such a mechanism has been postulated for the homologous protein PurR [31] and could potentially result in a significant elongation of the protein structure. Alternatively, the juxtapositions between the linkers, DNA, and regulatory domains may be distorted, resulting in changed interfaces and lower affinity for DNA [5]. If so, the dimensions of the ternary complex would not change much. These possibilities might be discriminated using small-angle X-ray solution scattering.

The results presented here are from a series of small-angle X-ray scattering studies performed on the LacI tetramer and its dimeric variant R3 [8; 32] (Figure 1B). Experiments included the apo-proteins and their complexes with DNA in the presence and absence of IPTG. Data interpretation was greatly enhanced by combining information from solution scattering with the available crystal structures. Earlier attempts to study the LacI structure using X-ray [33; 34; 35] and neutron [36; 37; 38; 39] solution scattering were limited by the lack dimeric variants or crystal structures to aid in interpretation of the scattering data.

Our scattering experiments reveal several new insights into the conformational dynamics of LacI in solution. The average solution structure of native tetrameric LacI differs from the crystal structure and the scattering data indicate it is an ensemble that includes a significant population of the crystal “V” form as well as more open forms. The dimensions of the dimeric R3-*o*DNA complex are in excellent agreement with the crystal structure. Removing DNA from R3 results in a significant elongation of the protein that is consistent with unfolding of the hinge-helix connecting the regulatory and DNA binding domains, and a similar elongation is indicated for the LacI tetramer. In contrast, results for the ternary complexes of dimeric and tetrameric LacI bound to *o*DNA and IPTG do *not* show evidence of significant elongation, and thus an interface between the linker and regulatory domain appears to remain formed in the ternary complex.

RESULTS

Basic structural parameters for native LacI and R3 from the solution scattering data

Small-angle X-ray scattering data were acquired for the native LacI tetramer (Figure 2A) and R3 dimer (Figure 2B) as apo-proteins and in their complexes with DNA, with and without bound IPTG. For native LacI, both operator (*o*DNA) and non-specific (*n*DNA) DNA complexes were explored. The concentration dependence of the scattering data was evaluated for all complexes. For apo-proteins < 3 mg/ml and their DNA complexes <5–10 mg/ml, the scattering data indicate well-behaved, mono-disperse particles (see below). For LacI-IPTG complexes, protein aggregation at all concentrations (data not shown) prevented structural analysis. Complexes of LacI and *n*DNA were incompletely formed as evidenced by the normalized *I(0)* values (Table 1), indicating that the binding affinity of LacI for the *n*DNA sequence used in this study is weaker than the DNA affinity of the ternary complex (LacI-*o*DNA-IPTG).

The data in Figure 2 and Table 1 represent all forms for which monodisperse solutions were obtained as judged by several criteria: (1) the scattering profiles are well behaved with respect to having linear Guinier regions (inserts Figures 2A&B), (2) calculated Porod volumes [40] are in reasonable agreement with the expected particle volumes [41], and (3) the normalized *I(0)* values are consistent with each of the samples having the correct molecular masses and

stoichiometries within experimental uncertainty, except for the incomplete complexes with *n*DNA as noted above. In addition, the R_g value determined for the LacI tetramer is in excellent agreement with previously reported measurements under conditions that were independently established to represent mono-disperse samples [35].

Guinier plots for elongated or rod-shaped particles (Figure 3) for LacI and R3 apo-proteins have a linear region with a "roll-over" at low *Q* characteristic of elongated particles of finite length and a reasonably well-defined average radius of gyration of cross section, R_c. Table 2 summarizes the R_g and R_c values determined using GNOM[42] for all protein concentrations measured, with the lowest concentration data being the same as in Figure 2 and Table 1 and thus known to be from monodisperse solutions. For both apo-proteins, while R_g values increase with increasing protein concentration due to aggregation, the R_c values remain constant, as do the extrapolated zero angle intensity values normalized for protein concentration ($I(0)_c/c$, where *c* is in mg/ml). This latter parameter is proportional to the average mass per unit length of an elongated or rod-shaped particle [43]. Thus the higher order aggregates, or oligomers, formed by each of the apo-proteins involve end-to-end associations that preserve the average cross-sectional radius and mass-per-unit-length. Of note, the experimental R_c value for dimeric R3 agrees well with that predicted from the crystal structure coordinates for the homodimer, whereas the R_c value for the apo-LacI tetramer is intermediate between the crystal structure predicted values for the homodimer and tetramer.

P(r) profiles calculated from the scattering data in Figure 2 are given in Figures 4 and 5 (upper panels). Each *P(r)* fit is evaluated as excellent according to the criteria of GNOM [42]. Consistent with the Guinier analyses for rod-shaped particles and the crystal structures, each experimental *P(r)* profile is asymmetric.

R3 undergoes a significant elongation upon removal of oDNA

The experimental R_g and d_{max} values for the R3-*o*DNA complex are in excellent agreement with those calculated using the crystal structure coordinates for the homodimer (Table 1). The χ^2 value of 1.3 for the fit of the calculated scattering profile with the desmeared experimental scattering profile is reasonable given that 11 additional C-terminal residues are present in R3 compared to the 1LBG homodimer (Figure 1B).

Comparison of R_g and d_{max} values for apo-R3 and the protein component of the homodimer crystal structure indicates that R3 undergoes a significant elongation upon removal of DNA. The protein lengthens by ~20 Å and R_g increases 20% while R_c remains the same which means that the elongation must be accomplished without any dramatic rearrangement among the regulatory domains. These large changes are consistent with unfolding the 7 residue hinge-helix that connects the DBD and regulatory domain. The program BUNCH [44; 45] was used to model apo-R3 using the crystal structure coordinates of one protein homodimer with the hinge helix residues assigned as a flexible linker and adding the 11 missing C-terminal residues as another segment of unknown structure. In the resulting optimized structure (R_g = 34.04 Å, d_{max} = 110 Å, χ^2 = 1.02), the DBDs move away from the regulatory domains with the hinge-helix sequence extended (Figure 6). Multiple BUNCH runs resulted in similar structures with statistically indistinguishable fits, all with the hinge-helix residues extended, but in different relative positions and orientations of the DBDs. The fit of one of the BUNCH models to the desmeared experimental *I(Q)* vs *Q* data is shown in Figure 7.

IPTG binding to the R3-oDNA complex results in small domain rearrangements without any significant elongation

Although the affinities of LacI and R3 for DNA are reduced several orders of magnitude by IPTG binding[13; 14; 15], the ternary complex does form under the conditions of our scattering

experiments, as evidence by the normalized *I(0)* values. This result is expected because the *o*DNA, protein, and IPTG concentrations were significantly higher than K_d values of either DNA or IPTG measured when forming the ternary complex[a] (see Materials and Methods, Table 1). The scattering data show that IPTG binding to R3-*o*DNA gives rise to a small (3%) increase in R_g and no measurable change in d_{max} (Table 1). Inspection of the difference *P(r)* between R3 with and without bound IPTG (Figure 4, lower panel) reveals this change is due to a redistribution of vector lengths within the scattering particle; there is a small increase (~5%) in the proportion of longer vector lengths (>60 Å) and a corresponding decrease in shorter vector lengths (<60 Å). These changes were reproduced in repeated experiments using independent sample preparations (data not shown). They indicate small reorientations of domains and/or sub-domains with the complex that could include the kinds of movements observed for the regulatory subdomains in the LacI core tetramer crystal structure upon IPTG binding [3; 5]. Another possibility is that the DNA-DBD makes small movements with respect to the regulatory domain.

The average solution conformation of the native LacI-oDNA complex differs significantly from the crystal structure

In the low resolution crystal structure of the LacI tetramer complexed with *o*DNA [3], two homodimers oligomerize via an anti-parallel four-helix bundle to form a V-shaped structure (Figure 1A, orange). For each monomer, an unstructured sequence (Figure 1A, dotted oval) joins the regulatory domain to the helix of the tetramerization domain. A comparison of the scattering profile for the LacI-*o*DNA complex with that predicted from the crystal structure shows very poor agreement, $\chi^2 = 5.23$ (Figure 7, Table 1). The derived R_g and d_{max} values are larger in solution, by 4.6 Å and 25 Å respectively, and large differences also are apparent between the scattering profiles in the mid-Q region corresponding to distances in the range ~60–120 Å that are characteristic of the separation of the homodimers. These results suggest that the larger R_g value in solution is because, on average, the dimer of homodimers adopts a more open configuration than seen in the crystal form. Such an interpretation is consistent with the measured R_c value for the apo-LacI being intermediate between the values predicted from the crystal structure coordinates for the tetramer and homodimer (Table 2), assuming that a significant portion of highly extended structures have R_c values consistent with the homodimer cross-section.

The program SASREF [45] optimizes the positions and orientations of domains or subunits with respect to scattering data given points of potential flexibility. Points of flexibility in LacI –by which the homodimers could move with respect to each other – were specified within the sequence of amino acids that joins the regulatory and tetramerization domains (Figure 1A, dotted oval). Multiple SASREF runs with increasingly loosened distance constraints between the tetramerization and regulatory domains were used to explore the full range of potential flexibility. The resulting optimizations gave improved fits to the data, converging to a set of structures that have altered juxtapositions between dimers – the V-shape is twisted. These structures have R_g values of 45–46 Å, R_c values of ~23–24 Å, and the χ^2 values for the fits of their predicted scattering profiles with experiments are somewhat improved (as low as 1.96). Although these parameters have moved toward a better fit with experiment, significant differences remain, especially in the critical low- and mid-Q and regions that are sensitive to the overall shape of the average scattering particle.

Therefore, we considered the possibility that the unstructured connection between the regulatory and tetramerization domains allows the homodimers to sample various orientations

[a]We monitored *o*DNA binding for R3 in the presence of IPTG and confirmed that binding is diminished by 1000 orders of magnitude in the current buffer (data not shown)

with respect to each other and thus the solution form of LacI-*o*DNA is an ensemble of conformations. A series of model structures was generated in which the acute angle between the homodimers was opened, in steps of 20°, from that observed in the crystal structure (approximately 23°) to a fully extended configuration. Combinations of these structures were used to model the observed scattering profile and excellent fits to the experimental data were obtained, with χ^2 values < 1.0 (Table 3).

The scattering profile of the best-fit ensemble or "flexible" model obtained, as determined by the lowest χ^2 value, is depicted in Figure 7 and its associated R_g value is in excellent agreement with experiment (Table 1). Figure 8 illustrates the three structures, and their relative populations, used to generate the fit shown in Figure 7. Not surprisingly, equally good fits are obtained with different ensemble populations. For example, a two structure fit (crystal plus 180° open) is statistically indistinguishable from the one shown in Figures 7 and 8 (Table 3). Notably, all the best-fit ensemble models ($\chi^2 < 1$) share a number of common features, specifically a substantial proportion of the crystal structure (59–64% of the population) as well as structures with the angle between the homodimers at least 60° larger than that in the crystal form (Table 3). As scattering data reflect the time and ensemble average of the solution species, ensembles of distinct conformational states versus conformationally dynamic structures are indistinguishable. However, if LacI can adopt the crystal structure as well as open forms in solution, then states intermediate between the crystal and most extended forms are likely to be present in a dynamic ensemble. Such a dynamic ensemble might be expected to increasingly favor more compact forms that minimize the solvent-exposed surface area, which might also be the form that would more readily form favorable crystal contacts.

The R_g and d_{max} values obtained for apo-LacI in solution also are larger than those predicted from the crystal structure; 2Å in R_g and 15 Å in d_{max}. Two possible sources may give rise to these differences: (1) The ensemble of structures sampled by the apo-protein has a different population distribution; or (2) The hinge-helix unfolds upon releasing *o*DNA. The fact that R3 and LacI tetramer have similar affinity for DNA (Table 4) suggests that hinge helix unfolding does occur. In support of this conclusion, attempts to model the apo-LacI scattering data with an ensemble that allowed only flexibility about the tetramerization domain while retaining a folded hinge-helix were unsuccessful (χ^2 values all > 2, data not shown). More elaborate modeling, adding unfolded/flexible hinge helices, was not attempted because would be too many variables to be reliably determined by the scattering data.

IPTG binding to native LacI-oDNA induces a small conformational change, consistent with relatively subtle movements of the homodimers with respect to each other

Addition of IPTG to native LacI-oDNA complex, at concentrations that maintain the ternary complex, causes no change in the measured R_g value. A subtle redistribution of vector lengths in the P(r) function is observed in the P(r) difference plot (Figure 5A) indicating that IPTG binding results in a small (~6%) increase in shorter vector lengths (< 60 Å) and a corresponding decrease in longer vector lengths (> 60 Å). This redistribution is opposite to the direction to that observed for R3 (Figure 4). Thus, in native LacI, the effects of the motions within the homodimer appear to be masked by other changes. One possibility is that the average relative dispositions of the homodimers and their bound DNA molecules change upon binding IPTG, again facilitated by flexible connections to the tetramerization domain. The direction of the changes in P(r) suggest a tightening of the LacI tetramer structures upon IPTG binding, i.e. a decrease in the average angle between the homodimers. Such a change might by driven by changes at interfaces due to internal shifts consequent to IPTG binding or changes in the net surface charge or hydrophobicity.

The observed changes in $P(r)$ upon LacI binding non-specific DNA (LacI-*n*DNA complex) show similar, though not identical, trends in movement to those observed for oDNA (Figure

5B). In this case the complexes appear to have only partially formed, as judged by the smaller R_g and d_{max}, values and narrower P(r) profile compared to the complex with oDNA, smaller than expected Porod volumes, and values of $I(0)_{norm}/I(0)_{Lys}$ significantly less than 1.0. Nonetheless, the direction and relative amplitudes of the changes in $P(r)$ are comparable with fully complexed LacI-*o*DNA suggesting that the effects of IPTG binding on the relative dispositions of the homodimers are similar for both DNA complexes, as well as for residual apo-LacI present in the solution.

DISCUSSION

Protein function cannot adequately be related to structure unless information is known about the structures of all major species. In proteins with allosteric modulation of function, information is often lacking about the ternary complex [46]. Functionally important motions are also difficult to detect and may impede crystallization. NMR experiments can circumvent these problems, but this technique is limited by the size of the protein. Small-angle X-ray scattering provides an opportunity to monitor the dimensions of proteins and their complexes in solution, and further information can be extracted from solution scattering data when it is interpreted in the context of known structures.

For LacI, crystal structures are known for three of the functionally-relevant states (Figure 9). However, the DBD and linkers are missing from the electron density of any structure lacking DNA [3; 19; 47]. Other evidence indicates that missing density might be due to motions and/or unfolding of the linker hinge helix [27; 28; 29; 30; 48; 49]. Our scattering experiments show that the removal of DNA results in an elongation of R3, consistent with unfolding of the hinge-helix sequence linking the regulatory and DBD domain. While inferred from many other experiments, the solution scattering data are the strongest evidence obtained to date for an unfolded linker hinge helix in an intact apo-LacI homodimer.

We also explored the possibility that the hinge helices unfold in the ternary complex (LacI-*o*DNA-IPTG). Previous comparison of the crystal structures for DNA-bound and IPTG-bound LacI variants indicated conformational changes in the regulatory domain that have the potential to interrupt or rearrange the linker interface [5]. However, prior to this study, little structural data were available for the ternary complex. We compared solution scattering for apo-R3 and R3-*o*DNA to that of the ternary complex (R3-*o*DNA-IPTG). No elongation is observed for the latter. Thus, transmission of the IPTG-initiated regulatory signal apparently does *not* involve a dramatic unfolding of the hinge-helix region (Figure 9). Instead, induced reorientation of the regulatory subdomains must alter the interface between the regulatory domain and the linkers, ultimately affecting the interface between the DBD and DNA. Such a model is consistent with mutagenesis of an engineered homologue, which showed that linker-interface mutations can alter repression of either or both the low- and high-affinity complexes [50].

The ternary *o*DNA complex must occur *in vivo*, but is short-lived, since excess genomic DNA effectively competes to displace the repressor protein from the operator [16; 17; 18]. Therefore, another question is whether the structure of LacI in the *o*DNA ternary complex is distinct from the repressor structure in complex with *n*DNA. We attempted to compare the solution scattering of LacI variants bound to *n*DNA, however under the current experimental conditions, these complexes were incompletely formed. Thus, the affinity of LacI for the *n*DNA sequence used in this study must be weaker than the affinity for *o*DNA in the presence of IPTG. While we cannot extract significant structural information from the current solution scattering data, insight has been obtained from other techniques (Figure 9). NMR studies with truncated LacI (DBD and linkers only) show that the hinge helices are unfolded when bound to *n*DNA and that the DNA ligand is B-form [24], lacking a kink that arises from the hinge helix side chains intercalating into the minor groove (*e.g.*[5; 25]). Circular dichroism studies that monitored

changes in *o*DNA and *n*DNA upon binding to full-length LacI also indicate that the two DNA structures differ [51].

For LacI tetramer, scattering experiments indicate that LacI-*o*DNA exists in solution as an ensemble of structures that includes significant proportions of the crystal structure and extended forms. This result would account for the poor resolution of the tetrameric LacI-*o*DNA crystal structure (4.8 Å) and the improved resolution obtained for the dimeric structures in which motions between the homodimers of a tetramer are eliminated [5]. The scattering data are consistent with the interpretation of TEM negative stain images [52] and AFM results [53], as well as with FRET data [54] that suggest open and closed LacI conformations interacting with DNA. Two computational studies also suggested the potential for more open and more tightly closed LacI forms [55; 56], whereas a third model predicted retention of the V-shape. [57]

The orientations of the two dimers within a tetramer have profound implications for the looping that occurs when one repressor tetramer binds two operator DNA sites (e.g. [11; 58;59]). The primary operator LacI (O1) is just upstream of the genes it regulates. Two additional minor operators are present in the operon; one (O2) is about 400 base pairs downstream in the early part of the lacZ gene and the second is at the end of the LacI gene (O3) (*e.g.* [60; 61]). Simultaneous binding of two operators significantly enhances repression ([10; 60; 61; 62; 63; 64]). This “biology” would be nicely served by the flexible dimer of homodimers, allowing the DBDs to find their pair of operator sites which may have variable spacing and also aiding in maintaining a looped DNA structure bound to two operator sites even when DNA dynamics might drive dissociation. Ultimately, the spacing of the operator binding sites and geometry of the intervening DNA are likely to affect the conformation adopted by tetrameric LacI, as predicted by Olsen and colleagues [56].

In conclusion, solution scattering data provide evidence for all three potential motions that were predicted for the LacI tetramer (Figure 1, arrows). In addition, the data have elucidated novel features of structures important to the functional cycle of the LacI homodimer (Figure 9). Notably, the ternary complex is measurably different from either of the singly liganded forms -results show small motions between domains or subdomains but the hinge helices remain compact. For dimers within a tetramer, conformations that are more open than the V-shaped crystal form were detected for apo- and DNA-bound states. A slight tightening or compression of the tetramer upon IPTG binding also is indicated. Thus, motions between subunits, domains, and subdomains are each correlated with different LacI functional states.

MATERIALS AND METHODS

Sample Preparation

Although wild-type LacI is a homotetramer, each dimer is an intact structural unit capable of binding one *o*DNA with the same intrinsic affinity as tetramer [32; 65]. This functional unit is preserved in the tetramer structure, which is best described as a “dimer of homodimers”. To distinguish motions within homodimers from those that might occur between homodimers, experiments utilized both the wild-type LacI homotetramer and dimeric versions. Solution scattering experiments were attempted with two versions of the LacI dimer: “-11”, which has the tetramerization domain deleted [7], and “R3”, which has the tetramerization domain substituted with the GCN4 dimerization domain [8; 32]. The −11 variant is more similar to the protein used to create the 1EFA dimeric crystal structure [5]. However, even at high concentrations, our solution scattering data showed that −11 is in equilibrium between the monomeric and dimeric forms (data not shown), consistent with previous results at lower concentrations [65]. Fortunately, R3 is fully dimeric at all concentrations studied.

The native LacI and R3 proteins were constitutively expressed in *E. coli* Blim cells [66] in 2×YT medium. Proteins were purified similarly to [67] and [32]. Briefly, after a 20-hr growth, cells from 1 L were collected by centrifugation, resuspended in breaking buffer (200 mM Tris-HCl, pH 7.6, 200 mM KCl, 20 mM $Mg(OAc)_2$, and 5% glucose) supplemented with protease inhibitor (one mini-tablet for 12 L growth), DTT (0.3 mM), and lysozyme (50 mg per 12 L growth), and stored at −20°. To start the purification, the cell paste was thawed on ice, then DNAse I was added to digest the genomic DNA. After centrifugation, ammonium sulfate (~ 35%) was added to precipitate proteins.

For wild-type LacI, the pellet of ammonium sulfate precipitate was resuspended in 0.09 M KP. "KP" refers to potassium phosphate buffer, pH 7.6, 5% glucose, and 0.3 mM DTT and the designated molar concentration (0.09 M in this case, same convention used below) refers to the molar concentration of potassium phosphate. The resuspended pellet was dialyzed against 1L of 0.09 M KP with three buffer changes. After centrifugation to remove particulate matter, the supernatant was loaded onto a phosphocellulose column (pre-equilibrated with 0.09 M KP). The column was then washed with 0.09 M KP and subsequently 0.12 M KP until the absorbance reached baseline values. The native LacI tetramer was eluted from the column using a gradient composed of 0.12 and 0.3 M KP. For R3, the ammonium sulfate pellet was resuspended in 0.05 M KP, dialyzed against 0.05 M KP, and loaded onto a phosphocellulose column, which was pre-equilibrated using 0.05 M KP. After washing extensively with 0.05 M KP to baseline absorbance, the dimeric proteins were eluted by 0.12 M KP.

Further purification of both proteins utilized ion exchange and size exclusion chromatography. The proteins were dialyzed into heparin buffer A (50 mM Tris-HCl, pH 7.4, 50 mM KCl, 1 mM EDTA, and 0.3 mM DTT), concentrated to a final volume of ~ 2ml using VIVA SPIN 20 (VIVA science, Goettingen, DE), and loaded onto a HiTrap heparin column (5 ml, GE) on an FPLC system. The proteins were eluted using a gradient of heparin buffer A and heparin buffer B (50 mM Tris-HCl, pH 7.4, 500 mM KCl, 1 mM EDTA, and 0.3 mM DTT). The eluate containing the desired protein elution was further dialyzed into Sepharose S200 buffer (100 mM Tris-HCl, pH 7.4, 100 mM KCl, 1 mM EDTA, and 0.3 mM DTT), concentrated using VIVA SPIN 20, and loaded onto a Sepharose S200 column for size-based purification. The fractions containing the target protein were collected and concentrated to desirable concentrations using VIVA SPIN 20.

DNA-binding affinities and operator release experiments with 40mer *o*DNA were determined similarly to [67]. Buffer compositions are noted in the footnotes to Table 4. As previously published in [32], K_d for *o*DNA and $[IPTG]_{mid}$ for R3 are very close to those of wild-type LacI. We attempted to determine binding to *o*DNA in the presence of IPTG for R3 and wild-type LacI. Conditions for these measurements are near the technical limit of filter binding experiments (high protein concentrations, weak binding affinities), but results do not suggest any difference between the two proteins (data not shown). In addition, operator release was performed simultaneously for R3 and wild-type LacI, with the same radio-labeled DNA and nitrocellulose filter. The resulting data showed very similar maximum and background counts, corresponding to very similar absolute quantities of bound protein across the range of IPTG concentrations (data not shown). Operator release experiments measure the allosteric change in the protein-DNA interaction upon binding IPTG[68]. Thus, we concluded that R3 is a good model of the overall allosteric conformational change that occurs within a single dimer of the LacI tetramer.

For solution scattering experiments, the sequence of the natural operator O^1 were incorporated into 21mer *o*DNA (5′-AATTGTGAGCGGATAACAATT-3′; Integrated DNA Technologies, Coralville, IA). Single-stranded oligomers were hybridized in polynucleotide kinase buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6). The annealing of operator was confirmed

by native polyacrymide gels. Oligomers for *n*DNA (5′-AGACATGCCTAGACATGCCTT-3′) [69] were treated similarly.

Samples for scattering experiments

Protein-DNA or protein-DNA-IPTG ternary complexes were prepared by mixing the required volumes of the protein, DNA, and IPTG stock solutions to produce the desired stoichiometries. The IPTG stock solution was made using the filtrates from concentrating the corresponding protein solution. The final IPTG concentration was 0.1–0.2 mM., which is above the K_d for IPTG in either the absence (1.5×10^{-6} M) or presence (2.5×10^{-5} M) of *o*DNA [70]. The K_d for R3 binding to IPTG in either the absence or presence of *o*DNA is within two-fold of wild-type LacI values[32]. The final concentration of each complex was also adjusted using the protein filtrate, and the filtrate served as an exact solvent blank in the scattering experiments. Solutions of binary and ternary complexes were routinely incubated for 30 minutes at room temperature before measurement.

The need for DTT in the samples prohibited accurate concentration determination by UV extinction. Instead, quantitative amino acid analysis was performed by Denis Winge at the University of Utah using an L-8800 Hitachi AAA instrument. Protein concentrations were calculated as the average of at least 3 runs using the known protein sequences and data for the nine amino acids with good separation of the peaks in the amino acid analysis chromatogram (Asp+Asn, Glu+Gln, Ala, Val, Ile, Leu, Phe, Lys, and Arg). Estimated errors in the concentrations based on repeated amino acid analyses are 10%. The concentrations of DNA were determined prior to annealing from the absorbance at 260 nm using known molar absorption coefficients for single-stranded DNA (204,100 L/(mole•cm) and assuming 100% yield in annealing. In preparing complexes and dilutions, all aliquots were measured by weight using a calibrated microbalance (four significant figures). The strong UV extinction coefficient for lysozyme allows more accurate concentration determination (est. 5%, extinction coefficient of 2.64 for 1 mg/ml/cm). Analysis of the forward scattering, *I(0)*, values normalized by protein concentration in mg/ml was used to check that the scattering particles were monodisperse and consistent with the expected stoichiometry (see below).

Scattering Data Acquisition and Analysis

X-ray scattering data were collected at 11°C using the small-angle instrument now at the University of Utah that has been described previously [71]. Scattering data were reduced to $I(Q)$ versus Q and analyzed as previously described [71]. $I(Q)$ is the scattered X-ray intensity per unit solid angle and Q is the amplitude of the scattering vector, given by $4\pi(\sin\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the scattered X-rays (1.54 Å). Scattering data were measured for the Q-range 0.01 – 0.3 $Å^{-1}$. The net scattering from the proteins and their complexes was determined by subtracting a normalized solvent blank scattering profile measured in the same capillary immediately before or after the sample measurements. Scattering data were collected for a range of protein concentrations for LacI, R3, and their complexes, in order to assess potential aggregation and/or inter-particle interference effects.

The $P(r)$ analyses were performed using the linear regularization method encoded in the program GNOM [42] and corrections for the slit geometry of the scattering instrument. $P(r)$ is the frequency of vector lengths connecting small-volume elements within the entire volume of the scattering particle, weighted by their scattering contrast. For a uniform scattering density object $P(r)$ goes to zero at the maximum dimension, d_{max}, of the object. Radius of gyration values (Rg) were calculated as the second moment of $P(r)$. Rg is the root-mean-square distance of all elemental volumes from the center-of-mass of the particle, weighted by their scattering contrast. In the case of the protein-DNA complexes studied here, electron dense DNA components are localized at one end of the structure. Therefore, the contrast weighting in these

parameters gives rise to smaller R_g and d_{max} values than would be the case for the equivalent uniform density object.

Estimates of the radius of gyration of cross section, R_c, and extrapolated $I(0)_c$ values were calculated using GNOM analysis for cross-sectional parameters[42] for rod-shaped particles. R_c is the weighted, root-mean-square distance of all elemental areas from the center of the cross-sectional area of a rod-shaped particle and $I(0)_c$ normalized to protein concentration in mg/ml is proportional to mass-per-unit length[43; 72].

Estimates for the volume of the scattering particles were obtained using the Porod invariant [40] and compared to calculated particle volumes, V_{calc} that were determined using the relevant amino acid and DNA sequences with standard atomic volumes for nucleic acids and proteins based on Voronoi volume calculations [41].

The accurate interpretation of scattering data in terms of structure requires dilute solutions that are mono-disperse. For proteins in solution, because their average electron densities are the same, *I(0)* is proportional simply to protein concentration (*c* in mg/ml) and molecular mass (*M*) [73]. Thus *I(0)*/*M*/*c* should be a constant for most proteins, and the ratio of this parameter for our proteins of interest against that for our scattering standard lysozyme is 1.0 if the protein molecules are monodisperse. The predicted *I(0)* values for protein-DNA complexes has to take into account the effect of the higher DNA contrast for X-ray scattering, and were calculated using the relationship:

$$I(0) = I(0)^N c(M_P \Delta\rho_P v_P + M_{DNA} \Delta\rho_{DNA} v_{DNA})^2$$

where $I(0)^N$ is the *I(0)* value for the lysozyme standard at infinite dilution and normalized to the square of its molecular weight, *c* is the molar concentration of the complex; M_P and M_{DNA} are the molecular weights in kDa of the protein and DNA; v_P and v_{DNA} are the partial specific volumes of the protein and DNA in cm^3/g; and $\Delta\rho_P$ and $\Delta\rho_{DNA}$ are the contrast values for the protein and DNA, calculated as the difference between electron density of the protein or DNA, ρ_P or ρ_{DNA}, and the electron density of the solvent ρ_s which is 0.33455 e/Å^3 for diluted aqueous buffered solutions. The electron densities for protein or DNA are determined using $\rho = N_e/v$, where N_e is an electron concentration in e/g, calculated as Z/M where Z is the total number of electrons in the protein or DNA and M the respective molecular weight. Partial specific volumes, v, for the proteins and DNA were determined based on the corresponding amino acid and DNA sequences, with standard atomic volumes for nucleic acids and proteins calculated using the Voronoi approximation [41]. The sensitivity of the partial specific volume for DNA to solution conditions results in additional 5% uncertainty in the calculation of predicted *I(0)* values for the protein-DNA complexes that is not accounted for in the error estimates given for the parameters in Table 1

Modeling

The program CRYSOL [74] was used to predict scattering profiles for the LacI tetramer from crystal structure coordinates 1LBG [3] from the Protein Data Bank. Homodimer parameters were calculated using coordinates of one homodimer from 1LBG (chains AB or ABEF for the apo-protein and *o*DNA complex, respectively). The program BUNCH [44; 45] was used to model the apo-R3 structure with a flexible hinge-helix region (residues 50–57).

The program SASREF [45] was used to find the best-fit model to the scattering data for the native LacI tetramer using the 1LBG crystal structure coordinates and allowing the homodimers the freedom to rotate and twist about the tetramerization domain while maintaining *P2* symmetry. The crystal structure coordinates for each homodimer were divided into two segments, one containing the two 24 C-terminal protein sequence segments from each monomer to form half of the tetramerization domain, and a second containing the remaining

protein coordinates for the homodimer plus those of their bound *o*DNA. The constraints used for the SASREF6 calculations loosely maintained the connectivities between the tetramerization domain and the regulatory domain of the homodimer (a total of two connections for each homodimer at opposite ends of the four helix bundle that forms the tetramerization domain) while allowing for freedom in the orientation of the domains with respect to each other.

An error weighted least squares fitting routine was used to determine the coefficients for the linear combination of scattering profiles based on the LacI-*o*DNA crystal structure (PDB 1LBG) and a series of "opened" structures created using the program COOT [75] and rotating, in 20° steps, one of the homodimers about the long axis through the middle of the tetramerization domain. The angle between the homodimers in the crystal structure is ~23°.

Acknowledgements

We thank Professor Clare Woodward for many helpful discussions throughout the course of these studies and for critically reading this manuscript. JT was supported by an ARC Federation Fellowship (FF0457488) and the scattering experiments were performed at the University of Utah supported by U. S. Department of Energy, Grant No. DE-FG02-05ER64026. LSK was supported by NIH P20 RR17708 from the Institutional Development Award program of the National Center for Research Resources and GM079423. KSM was supported by NIH GM22441 and Robert A. Welch Foundation C-576.

References

1. Ogata RT, Gilbert W. An amino-terminal fragment of lac repressor binds specifically to lac operator. Proc Natl Acad Sci U S A 1978;75:5851–5854. [PubMed: 16592594]
2. Platt T, Files JG, Weber K. Lac repressor. Specific proteolytic destruction of the NH 2 -terminal region and loss of the deoxyribonucleic acid-binding activity. J Biol Chem 1973;248:110–21. [PubMed: 4571224]
3. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science 1996;271:1247–54. [PubMed: 8638105]
4. Bell CE, Lewis M. Crystallographic analysis of Lac repressor bound to natural operator O1. J Mol Biol 2001;312:921–6.
5. Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol 2000;7:209–14. [PubMed: 10700279]
6. Weickert MJ, Adhya S. A family of bacterial regulators homologous to Gal and Lac repressors. J Biol Chem 1992;267:15869–74. [PubMed: 1639817]
7. Chen J, Matthews KS. Deletion of lactose repressor carboxyl-terminal domain affects tetramer formation. J Biol Chem 1992;267:13843–50. [PubMed: 1629185]
8. Alberti S, Oehler S, von Wilcken-Bergmann B, Müller-Hill B. Genetic analysis of the leucine heptad repeats of Lac repressor: evidence for a 4-helical bundle. Embo J 1993;12:3227–36. [PubMed: 8344260]
9. Alberti S, Oehler S, von Wilcken-Bergmann B, Krämer H, Müller-Hill B. Dimer-to-tetramer assembly of *lac* repressor involves a leucine heptad repeat. New Biologist 1991;3:57–62.
10. Oehler S, Eismann ER, Kramer H, Müller-Hill B. The three operators of the lac operon cooperate in repression. Embo J 1990;9:973–9. [PubMed: 2182324]
11. Matthews KS. DNA Looping. Microbiological Reviews 1992;56:123–136. [PubMed: 1579106]
12. Gilbert W, Müller-Hill B. Isolation of the Lac Repressor. Proc Natl Acad Sci U S A 1966;56:1891–1898. [PubMed: 16591435]
13. Riggs AD, Newby RF, Bourgeois S. lac repressor--operator interaction. II. Effect of galactosides and other ligands. J Mol Biol 1970;51:303–14. [PubMed: 4320936]
14. Jobe A, Bourgeois S. lac Repressor-operator interaction. VI. The natural inducer of the lac operon. J Mol Biol 1972;69:397–408. [PubMed: 4562709]

15. Barkley MD, Riggs AD, Jobe A, Burgeois S. Interaction of effecting ligands with lac repressor and repressor-operator complex. Biochemistry 1975;14:1700–12. [PubMed: 235964]
16. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. Science 2007;316:1191–4. [PubMed: 17525339]
17. Lin S, Riggs AD. A comparison of lac repressor binding to operator and to nonoperator DNA. Biochem Biophys Res Commun 1975;62:704–10. [PubMed: 1120077]
18. Lin S, Riggs AD. The general affinity of lac repressor for E. coli DNA: implications for gene regulation in procaryotes and eucaryotes. Cell 1975;4:107–11. [PubMed: 1092468]
19. Daber R, Stayrook S, Rosenberg A, Lewis M. Structural analysis of lac repressor bound to allosteric effectors. J Mol Biol 2007;370:609–19. [PubMed: 17543986]
20. Friedman AM, Fischmann TO, Steitz TA. Crystal structure of lac repressor core tetramer and its implications for DNA looping. Science 1995;268:1721–7. [PubMed: 7792597]
21. Chuprina VP, Rullmann JA, Lamerichs RM, van Boom JH, Boelens R, Kaptein R. Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. J Mol Biol 1993;234:446–62. [PubMed: 8230225]
22. Slijper M, Bonvin AM, Boelens R, Kaptein R. Refined structure of lac repressor headpiece (1–56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the lac operator. J Mol Biol 1996;259:761–73. [PubMed: 8683581]
23. Kalodimos CG, Bonvin AM, Salinas RK, Wechselberger R, Boelens R, Kaptein R. Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain. Embo J 2002;21:2866–76. [PubMed: 12065400]
24. Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, Kaptein R. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 2004;305:386–9. [PubMed: 15256668]
25. Spronk CA, Bonvin AM, Radha PK, Melacini G, Boelens R, Kaptein R. The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. Structure 1999;7:1483–92. [PubMed: 10647179]
26. Sams CF, Vyas NK, Quiocho FA, Matthews KS. Predicted structure of the sugar-binding site of the lac repressor. Nature 1984;310:429–30. [PubMed: 6462229]
27. Ha JH, Spolar RS, Record MT Jr. Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. J Mol Biol 1989;209:801–16. [PubMed: 2585510]
28. Spronk CAEM, Slijper M, van Boom JH, Kaptein R, Boelens R. Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. Nature Structural Biology 1996;3:916–919.
29. Kalodimos CG, Folkers GE, Boelens R, Kaptein R. Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. Proc Natl Acad Sci U S A 2001;98:6039–44. [PubMed: 11353825]
30. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. Science 1994;263:777–784. [PubMed: 8303294]
31. Schumacher MA, Choi KY, Lu F, Zalkin H, Brennan RG. Mechanism of corepressor-mediated specific DNA binding by the purine repressor. Cell 1995;83:147–55. [PubMed: 7553867]
32. Chen J, Alberti S, Matthews KS. Wild-type operator binding and altered cooperativity for inducer binding of lac repressor dimer mutant R3. J Biol Chem 1994;269:12482–7. [PubMed: 8175655]
33. Culard F, Charlier M, Maurizot JC, Tardieu A. Lac repressor-Lac operator complexes. Solution X-ray scattering and electrophoretic studies. Eur Biophys J 1987;14:169–78. [PubMed: 3549272]
34. Pilz I, Goral K, Kratky O, Bray RP, Wade-Jardetzky NG, Jardetzky O. Small-angle X-ray studies of the quaternary structure of the lac repressor from Escherichia coli. Biochemistry 1980;19:4087–90. [PubMed: 6996716]
35. McKay DB, Pickover CA, Steitz TA. Escherichia coli lac repressor is elongated with its operator DNA binding domains located at both ends. J Mol Biol 1982;156:175–83. [PubMed: 7047750]
36. Charlier M, Maurizot JC, Zaccai G. Nonspecific binding of lac repressor to DNA. II. A small-angle neutron-scattering study. Biophys Chem 1983;18:313–22. [PubMed: 6362732]

37. Charlier M, Maurizot JC, Zaccai G. Neutron-scattering studies of lac repressor: a low-resolution model. J Mol Biol 1981;153:177–82. [PubMed: 7338911]
38. Charlier M, Maurizot JC, Zaccai G. Neutron scattering studies of lac repressor. Nature 1980;286:423–5. [PubMed: 6772965]
39. Maurizot JC, Charlier M, Helene C. Lac repressor binding to poly (d(A-T)). Conformational changes. Biochem Biophys Res Commun 1974;60:951–7. [PubMed: 4429568]
40. Porod G. Die Roentgenkleinwinkel-Steuung von Dichtgepackten Kolloiden Systemen, I Teil. Kolloid Z Biol 1951;124:83–111.
41. Voss NR, Gerstein M. Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. J Mol Biol 2005;346:477–92. [PubMed: 15670598]
42. Svergun DI. Mathematical methods in small-angle scattering data analysis. J Appl Cryst 1991;24:485–492.
43. Hjelm RJ. The small-angle approximation of X-ray and neutron scatter from rigid rods of non-uniform cross section of finite length. Journal of Applied Crystallography 1985;18:452–460.
44. Petoukhov, MV.; Svergun, DI. http://www.emblhamburg.de/ExternalInfo/Research/Sax/bunch.html
45. Petoukhov MV, Svergun DI. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data 2005;89:1237–1250.
46. Reinhart GD. Quantitative analysis and interpretation of allosteric behavior. Methods Enzymol 2004;380:187–203. [PubMed: 15051338]
47. Bell CE, Barry J, Matthews KS, Lewis M. Structure of a variant of lac repressor with increased thermostability and decreased affinity for operator. J Mol Biol 2001;313:99–109. [PubMed: 11601849]
48. Swint-Kruse L, Matthews KS, Smith PE, Pettitt BM. Comparison of simulated and experimentally determined dynamics for a variant of the LacI DNA-binding domain, Nlac-P. Biophys J 1998;74:413–21. [PubMed: 9449341]
49. Swint-Kruse L, Larson C, Pettitt BM, Matthews KS. Fine-tuning function: correlation of hinge domain interactions with functional distinctions between LacI and PurR. Protein Sci 2002;11:778–94. [PubMed: 11910022]
50. Tungtur S, Egan SM, Swint-Kruse L. Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. Proteins: Structure, Function, and Bioinformatics 2007;68:375–388.
51. Culard F, Maurizot JC. Binding of lac repressor induces different conformational changes on operator and non-operator DNAs. FEBS Lett 1982;146:153–6. [PubMed: 6754436]
52. Ruben GC, Roos TB. Conformation of Lac repressor tetramer in solution, bound and unbound to operator DNA. Microsc Res Tech 1997;36:400–16. [PubMed: 9140942]
53. Virnik K, Lyubchenko YL, Karymov MA, Dahlgren P, Tolstorukov MY, Semsey S, Zhurkin VB, Adhya S. “Antiparallel” DNA loop in gal repressosome visualized by atomic force microscopy. J Mol Biol 2003;334:53–63. [PubMed: 14596799]
54. Edelman LM, Cheong R, Kahn JD. Fluorescence resonance energy transfer over approximately 130 basepairs in hyperstable lac repressor-DNA loops. Biophys J 2003;84:1131–45. [PubMed: 12547794]
55. Hawkins RJ, McLeish TC. Coarse-grained model of entropic allostery. Phys Rev Lett 2004;93:098104. [PubMed: 15447145]
56. Swigon D, Coleman BD, Olson WK. Modeling the Lac repressor-operator assembly: The influence of DNA looping on Lac repressor conformation. PNAS 2006;103:9879–9884. [PubMed: 16785444]
57. Villa E, Balaeff A, Schulten K. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. Proc Natl Acad Sci U S A 2005;102:6783–8. [PubMed: 15863616]
58. Eismann E, von Wilcken-Bergmann B, Müller-Hill B. Specific destruction of the second lac operator decreases repression of the lac operon in Escherichia coli fivefold. J Mol Biol 1987;195:949–52. [PubMed: 3116268]

59. Krämer H, Niemöller M, Amouyal M, Revet B, von Wilcken-Bergmann B, Müller-Hill B. lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. EMBO J 1987;6 (5):1481–1491. [PubMed: 3301328]
60. Reznikoff WS, Winter RB, Hurley CK. The location of the repressor binding sites in the lac operon. Proc Natl Acad Sci U S A 1974;71:2314–8. [PubMed: 4601586]
61. Pfahl M, Gulde V, Bourgeois S. “Second” and “third operator” of the lac operon: an investigation of their role in the regulatory mechanism. J Mol Biol 1979;127:339–44. [PubMed: 430569]
62. Besse M, von Wilcken-Bergmann B, Müller-Hill B. Synthetic *lac* operator mediates repression through *lac* repressor when introduced upstream and downstream from *lac* promoter. EMBO J 1986;5:1377–1381. [PubMed: 3015603]
63. Mossing MC, Record MT Jr. Upstream operators enhance repression of the lac promoter. Science 1986;233:889–92. [PubMed: 3090685]
64. Brenowitz M, Pickar A, Jamison E. Stability of a Lac repressor mediated “looped complex”. Biochemistry 1991;30:5986–98. [PubMed: 2043636]
65. Chen J, Matthews KS. Subunit dissociation affects DNA binding in a dimeric lac repressor produced by C-terminal deletion. Biochemistry 1994;33:8728–35. [PubMed: 8038163]
66. Wycuff DR, Matthews KS. Generation of an Ara-C-*ara*BAD promoter-regulated T7 expression system. Analytical Biochemistry 2000;277:67–73. [PubMed: 10610690]
67. Zhan H, Swint-Kruse L, Matthews KS. Extrinsic Interactions Dominate Helical Propensity in Coupled Binding and Folding of the Lactose Repressor Protein Hinge Helix. Biochemistry 2006;45:5896–5906. [PubMed: 16669632]
68. Swint-Kruse L, Zhan H, Matthews KS. Integrated insights from simulation, experiment, and mutational analysis yield new details of LacI function. Biochemistry 2005;44:11201–13. [PubMed: 16101304]
69. Falcon CM, Matthews KS. Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. Biochemistry 2000;39:11074–83. [PubMed: 10998245]
70. Daly TJ, Matthews KS. Allosteric regulation of inducer and operator binding to the lactose repressor. Biochemistry 1986;25:5479–84. [PubMed: 3535880]
71. Heidorn DB, Trewhella J. Comparison of the crystal and solution structures of calmodulin and troponin C. Biochemistry 1988;27:909–15. [PubMed: 3365370]
72. Guinier, A.; Fournet, G. Small-Angle Scattering of X-rays. John Wiley and Sons; New York: 1955. p. 128-129.
73. Krigbaum WR, Kugler FR. Molecular conformation of egg-white lysozyme and bovine alpha-lactalbumin in solution. Biochemistry 1970;9:1216–23. [PubMed: 5461510]
74. Svergun DI, Barberato C, Koch MHJ. CRYSOL — a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Cryst 1995;28:768–773.
75. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132. [PubMed: 15572765]
76. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Cryst 1992;25:495–503.
77. Spronk CA, Slijper M, van Boom JH, Kaptein R, Boelens R. Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. Nat Struct Biol 1996;3:916–9. [PubMed: 8901866]
78. Swint-Kruse L. Using networks to identify fine structural differences between functionally distinct protein states. Biochemistry 2004;43:10886–95. [PubMed: 15323549]

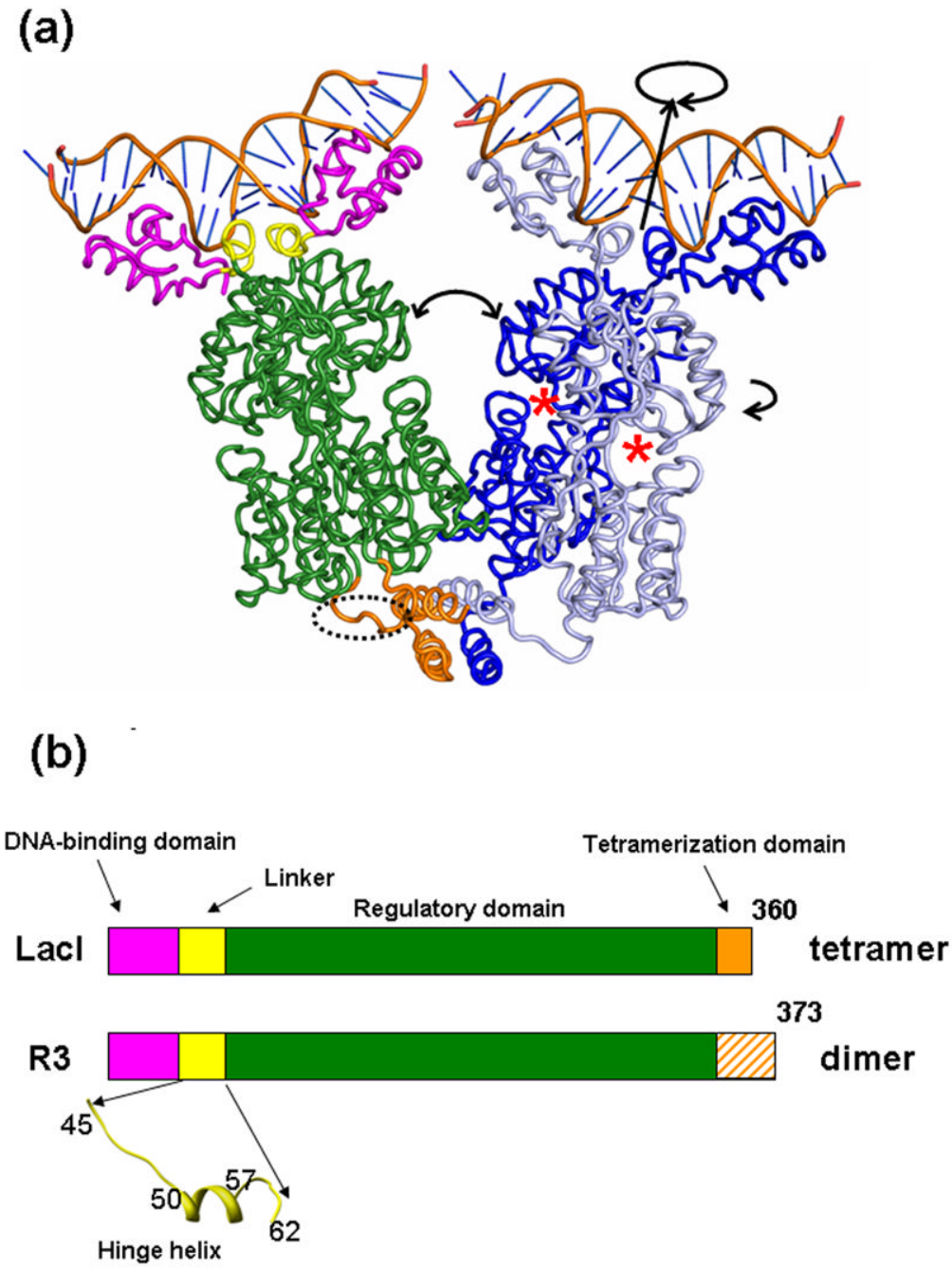
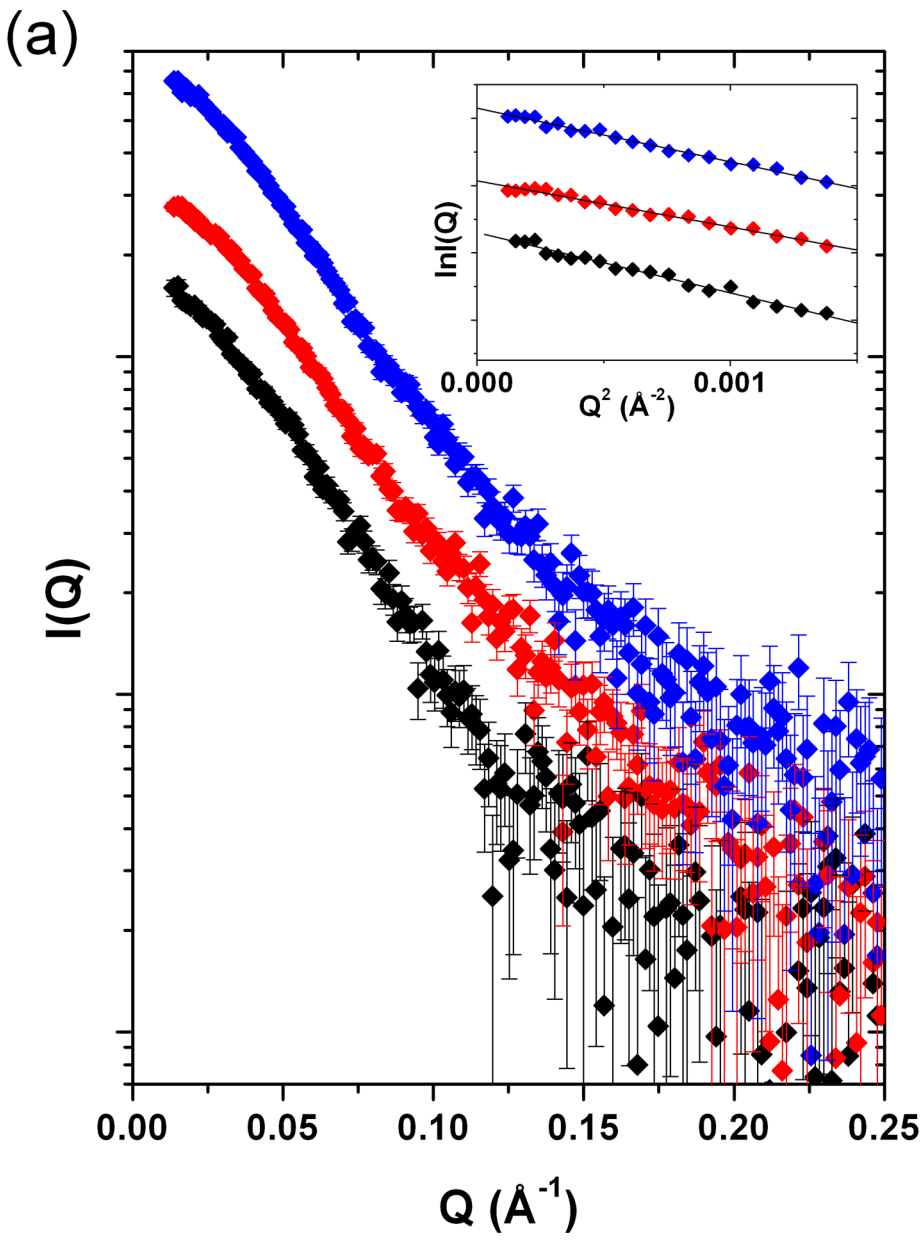


Figure 1. (a) Crystal structure of LacI tetramer bound to DNA and (b) schematic drawing showing the sequence differences among the LacI variants

Crystal structure coordinates are from the Protein Data Bank submission 1LBG[3]. Two bound DNA sequences are depicted at the top of the structure (orange ribbons with blue sticks for the base pairs). The protein is represented by a ribbon that traces the Cα coordinates. Color-coding of the protein structure corresponds to the schematic in (B); (magenta = DNA-binding domain, DBD; yellow = DBD-regulatory domain linker, including the hinge helix; green = regulatory/inducer-binding domain; orange = tetramerization domain). The right dimers are colored to highlight the individual monomers in light and dark blue. Potential motions between subunits, domains, and subdomains are indicated with black arrows. The two IPTG binding sites of one

homodimer are indicated with red asterisks (*). The flexible region that joins the regulatory domain to the tetramerization domain is indicated with a dotted oval for one monomer. R3 differs from wild-type LacI at the C-terminus; the wild-type LHR4 tetramerization sequence is substituted with the GCN4 dimerization domain (B, orange) [8].



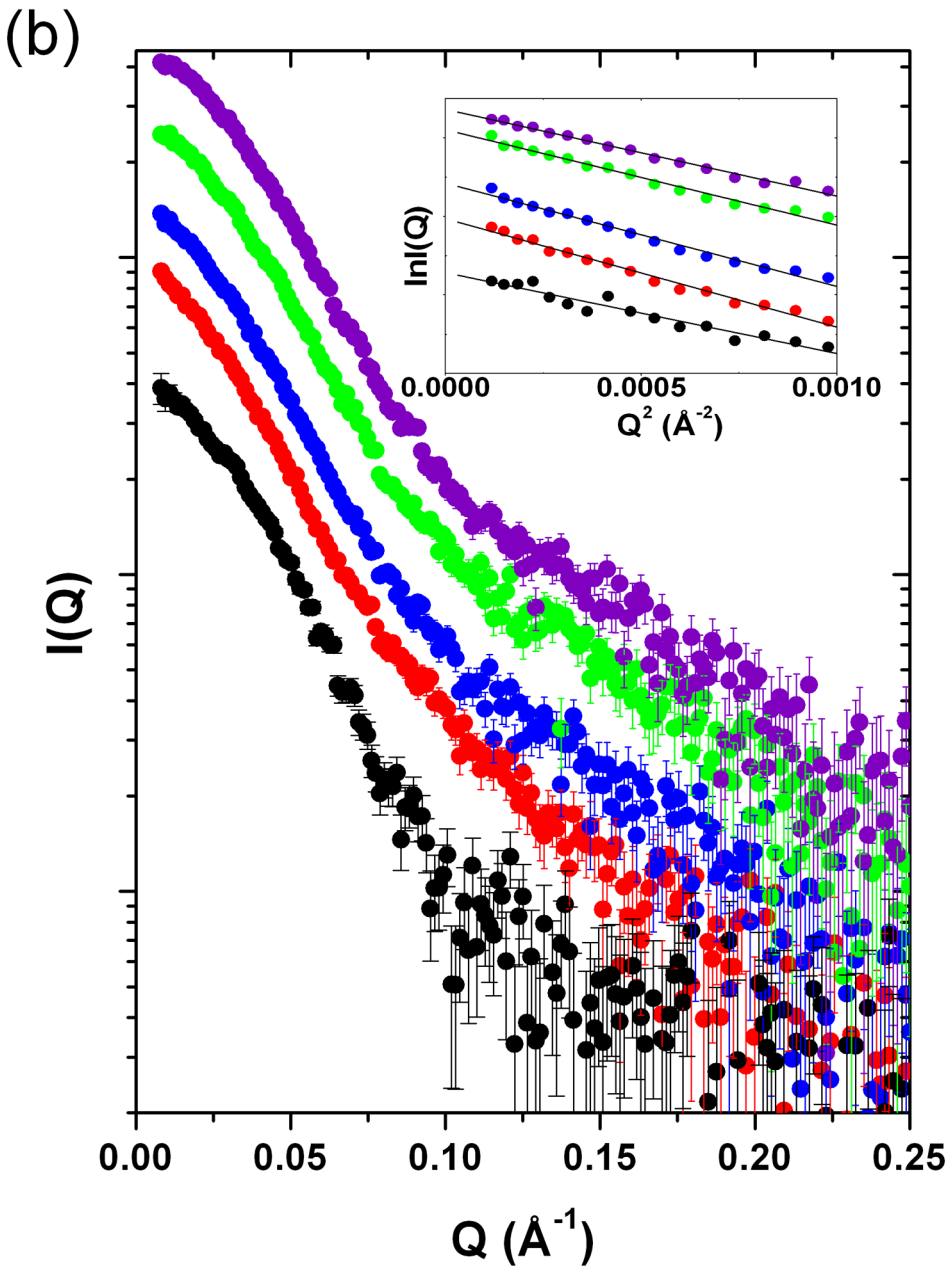


Figure 2. Scattering data for the native LacI tetramer, the dimer variants, and their complexes with DNA and IPTG

Data are shown as *I(Q)* versus *Q*, offset on the vertical scale for clarity. Guinier plots for the *Q*-range satisfying QR_g<1.3 for each data set are shown in the inserts. Error bars represent the propagated statistical errors. The key to the symbols is as follows: (a) ◆ apo-R3 dimer, ♦ R3-*o*DNA, ♦ R3-*o*DNA-IPTG. (b) ● apo-LacI tetramer, ● LacI-*o*DNA, ● LacI-*o*DNA-IPTG, ● LacI-*n*DNA, ● LacI-*n*DNA-IPTG;. Protein concentrations are given in Table 1.

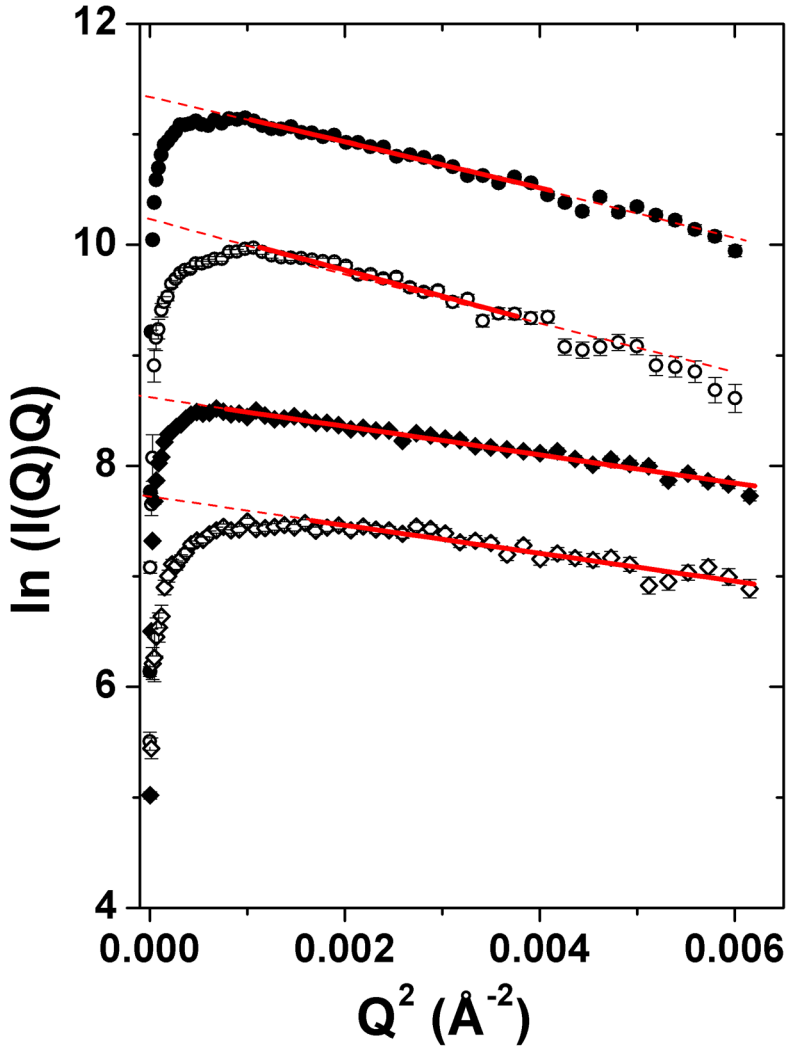
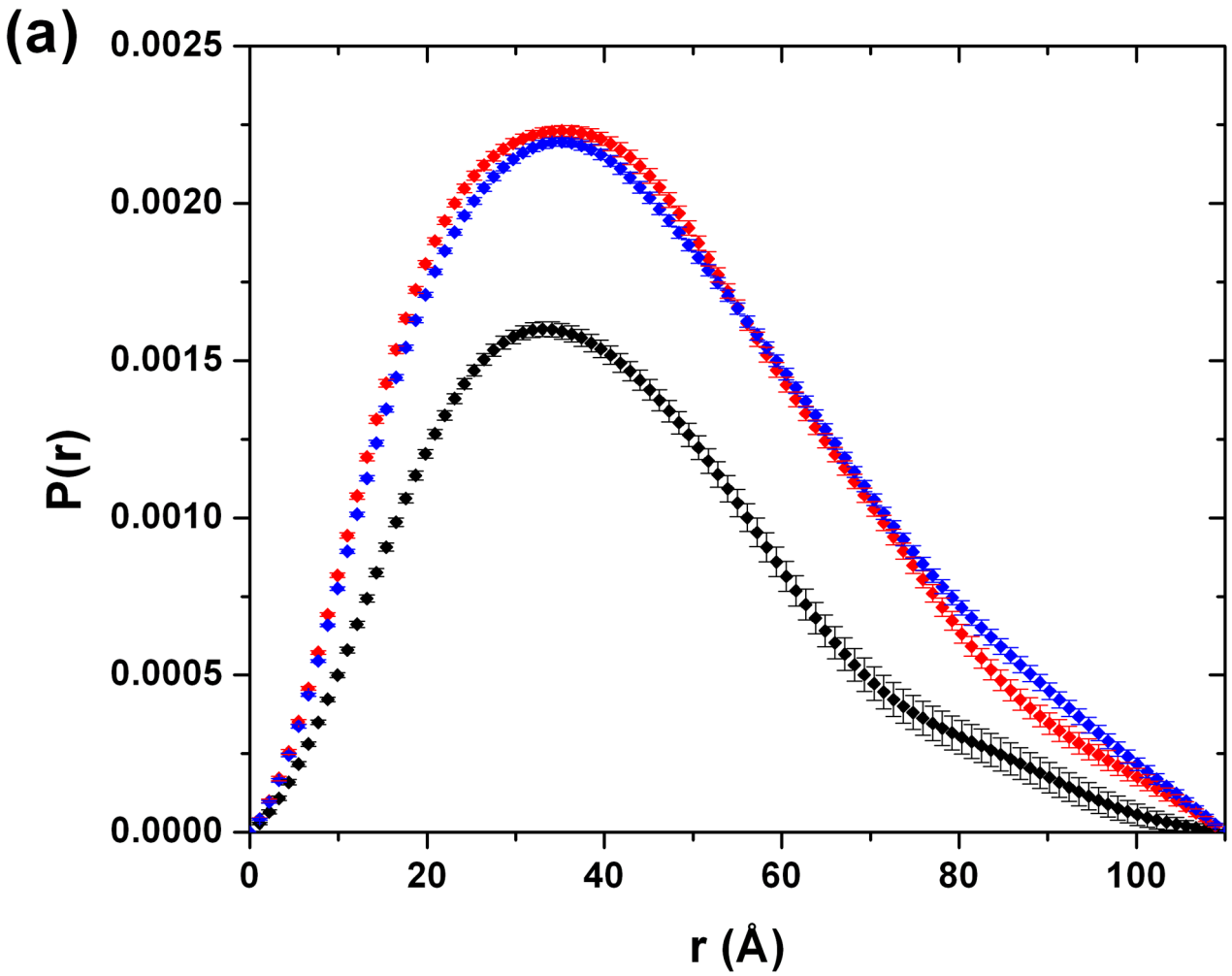


Figure 3. Guinier plots for rod-shaped particles

Solid lines indicate linear least-squares fits to the scattering data in the Guinier region, and dashed lines show the extrapolation of those fits. Plots are offset for clarity. Data are shown for the highest protein concentration samples measured and for samples known to be monodisperse (corresponding to the data shown in Figure 2 and Table 1). The key to the symbols is ● 28.1 mg/ml, ◦ 2.5 mg/ml apo-LacI tetramer; ◆ 21 mg/ml, ◇ 2.6 mg/ml apo-R3.



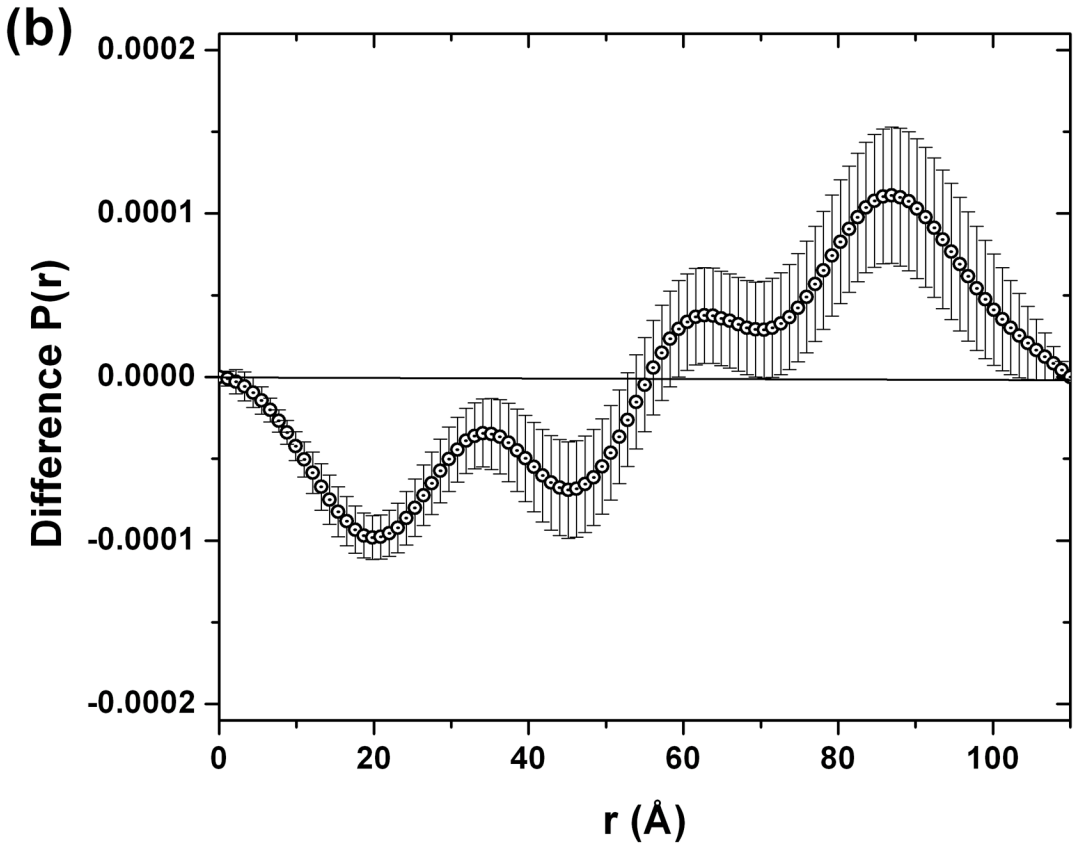
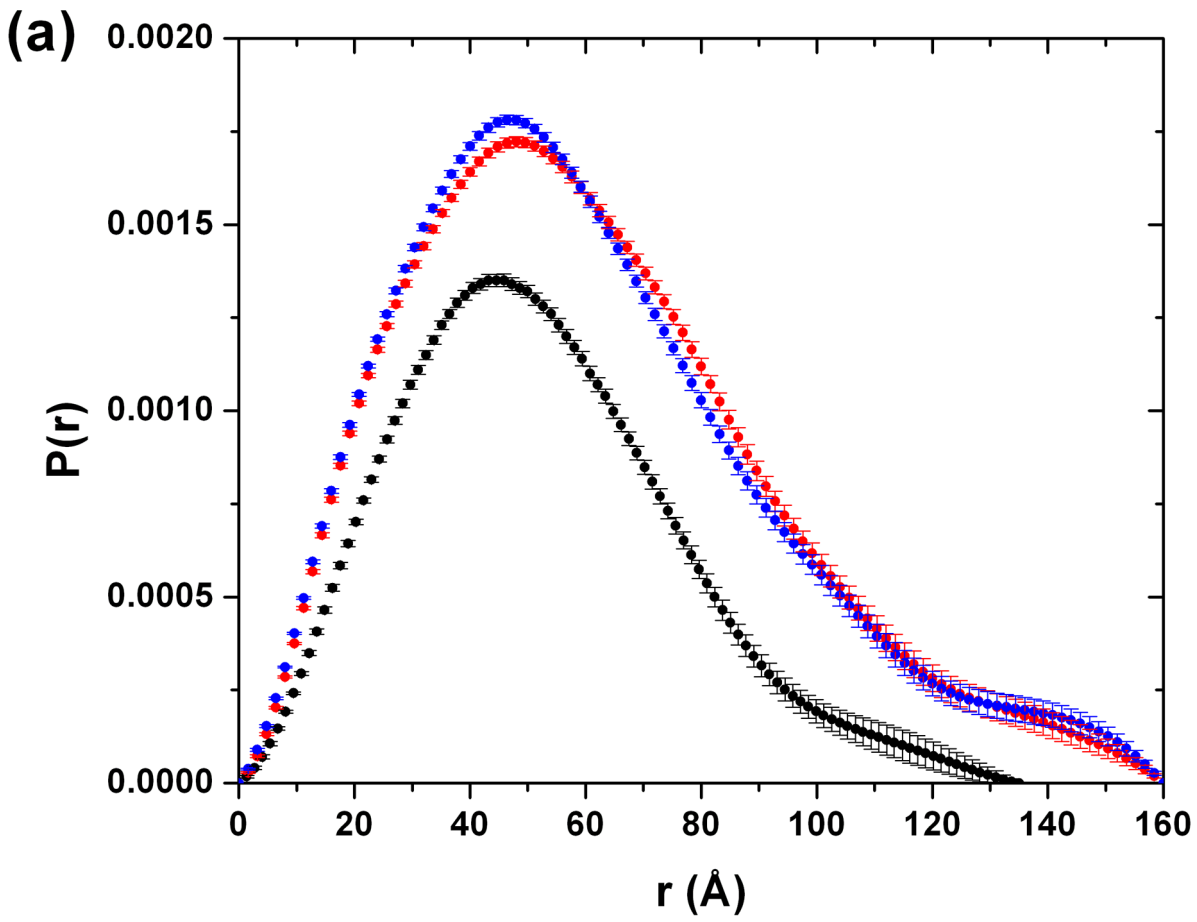
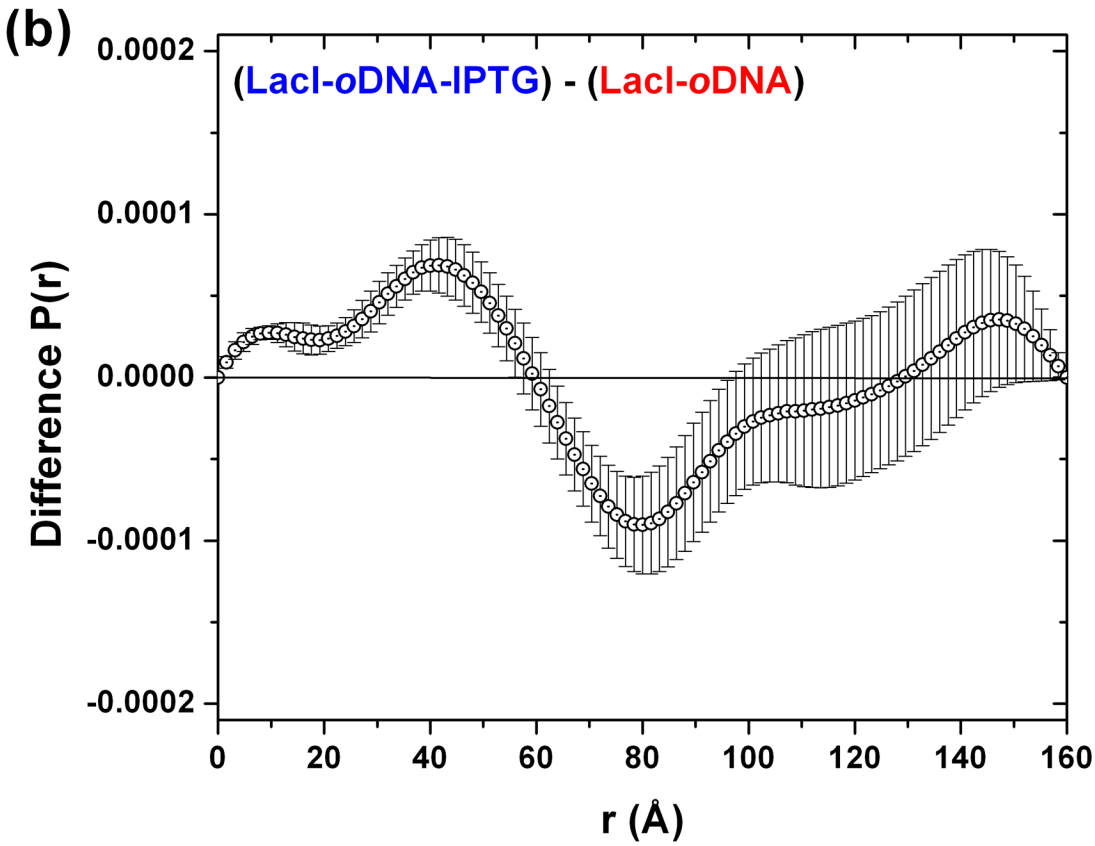
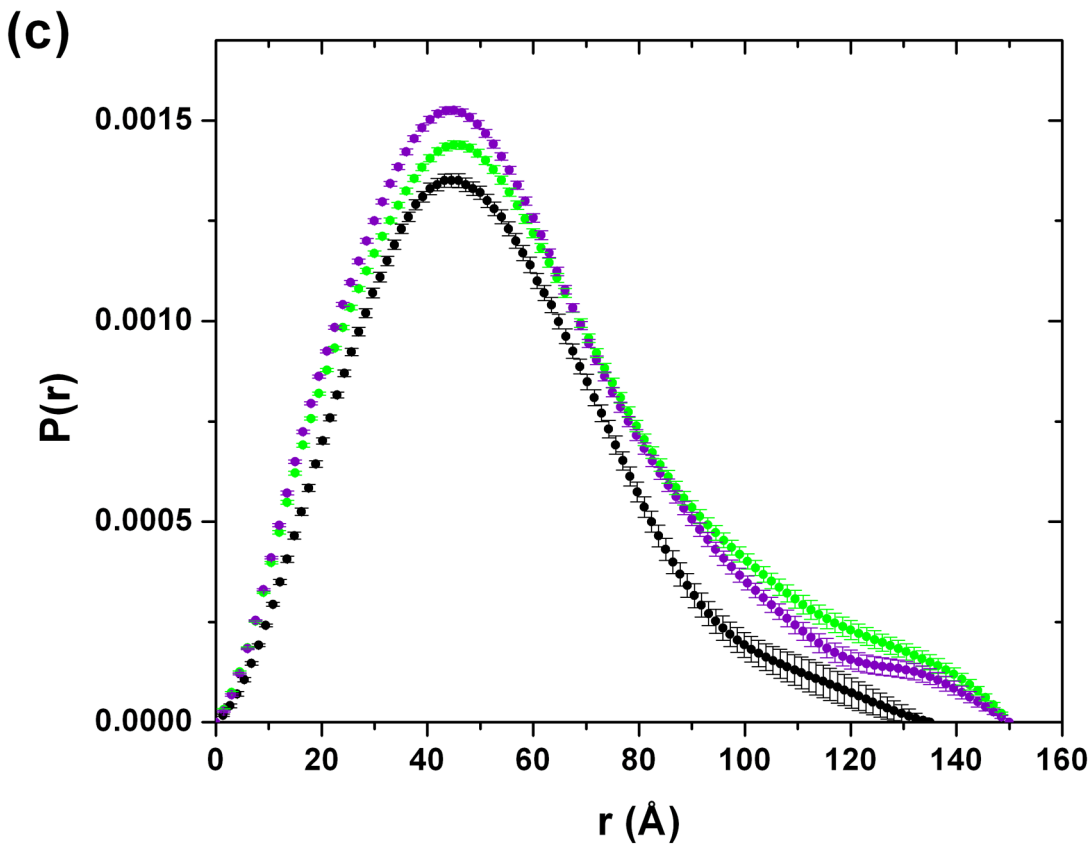


Figure 4. *P(r)* functions and difference *P(r)* plots for the R3 dimer and its complexes with DNA

(a) *P(r)* functions calculated form the data in Figure 2A for ◆ apo-R3 dimer, ♦ R3-*o*DNA, and ♦ R3-*o*DNA-IPTG. (b) Difference *P(r)* plots for (R3-*o*DNA-IPTG) – (R3-*o*DNA).







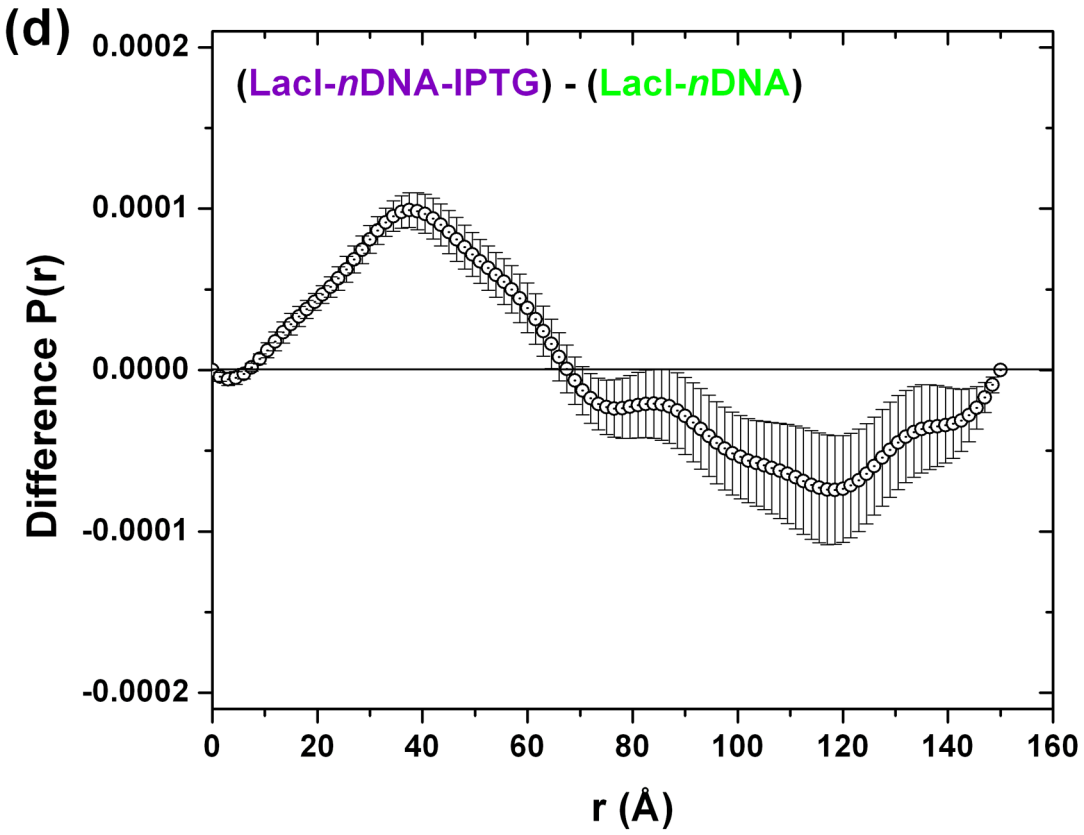


Figure 5. *P(r)* functions and difference *P(r)* plots for LacI tetramer and its complexes

Top panels (a & c) show the *P(r)* functions calculated from the scattering data in Figure 2B for: (a) ● LacI apo-protein, ● LacI-*o*DNA, and ● LacI-*o*DNA-IPTG complexes; and (c) ● LacI apo-protein, ● LacI-*n*DNA, and ● LacI-*n*DNA-IPTG. The bottom panels (b & d) show difference *P(r)* plots using the same convention as in Figure 4, (protein-DNA-IPTG) –(protein-DNA).

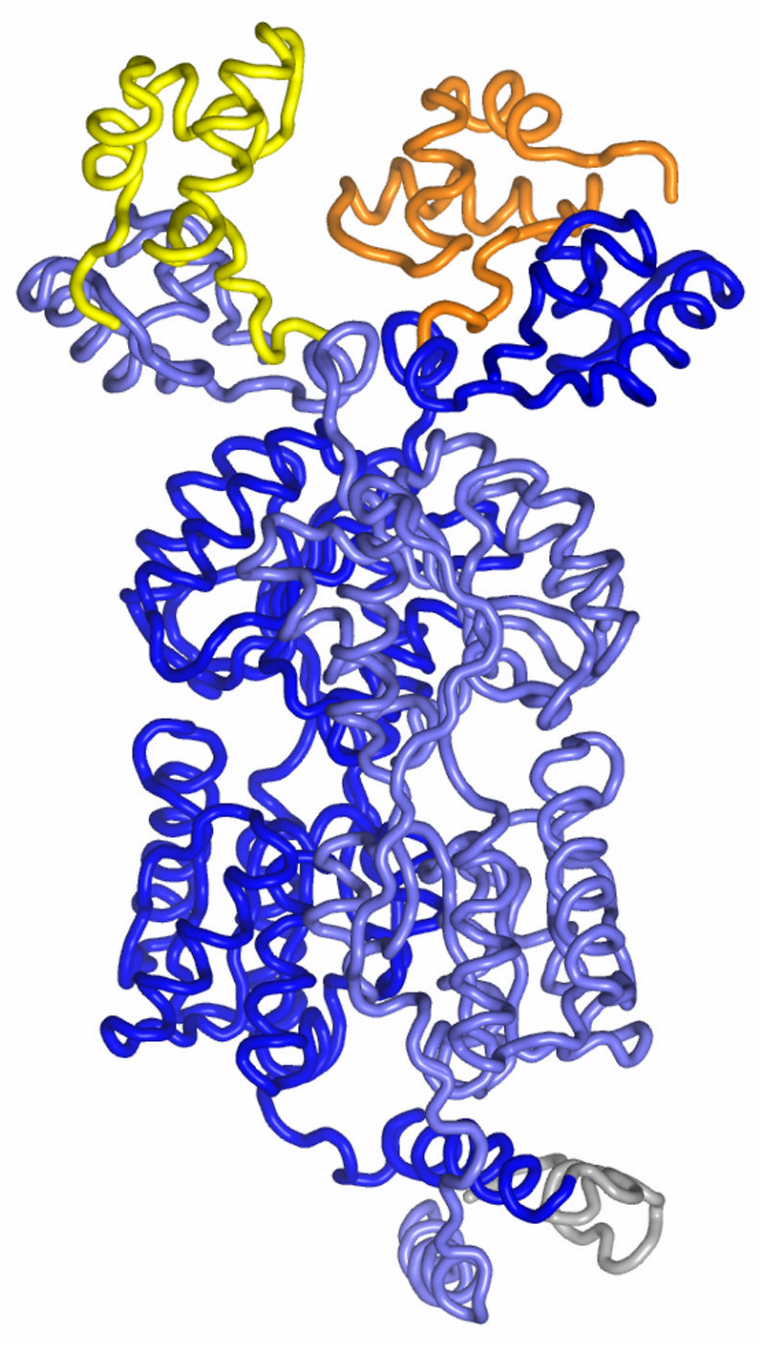


Figure 6. Comparison of the LacI homodimer from the crystal structure with the BUNCH-derived model for apo-R3

Superposed on the protein component of one LacI-homodimer form (chains A and B from 1LBG[3], blue and light blue) is the BUNCH-derived model obtained using the homodimer crystal structure coordinates and assigning the hinge-helices (residues 50-57) to arbitrary structure (resultant DBD and hinge-helices are yellow and orange) and adding an 11-residue C-terminus of unknown structure (gray).

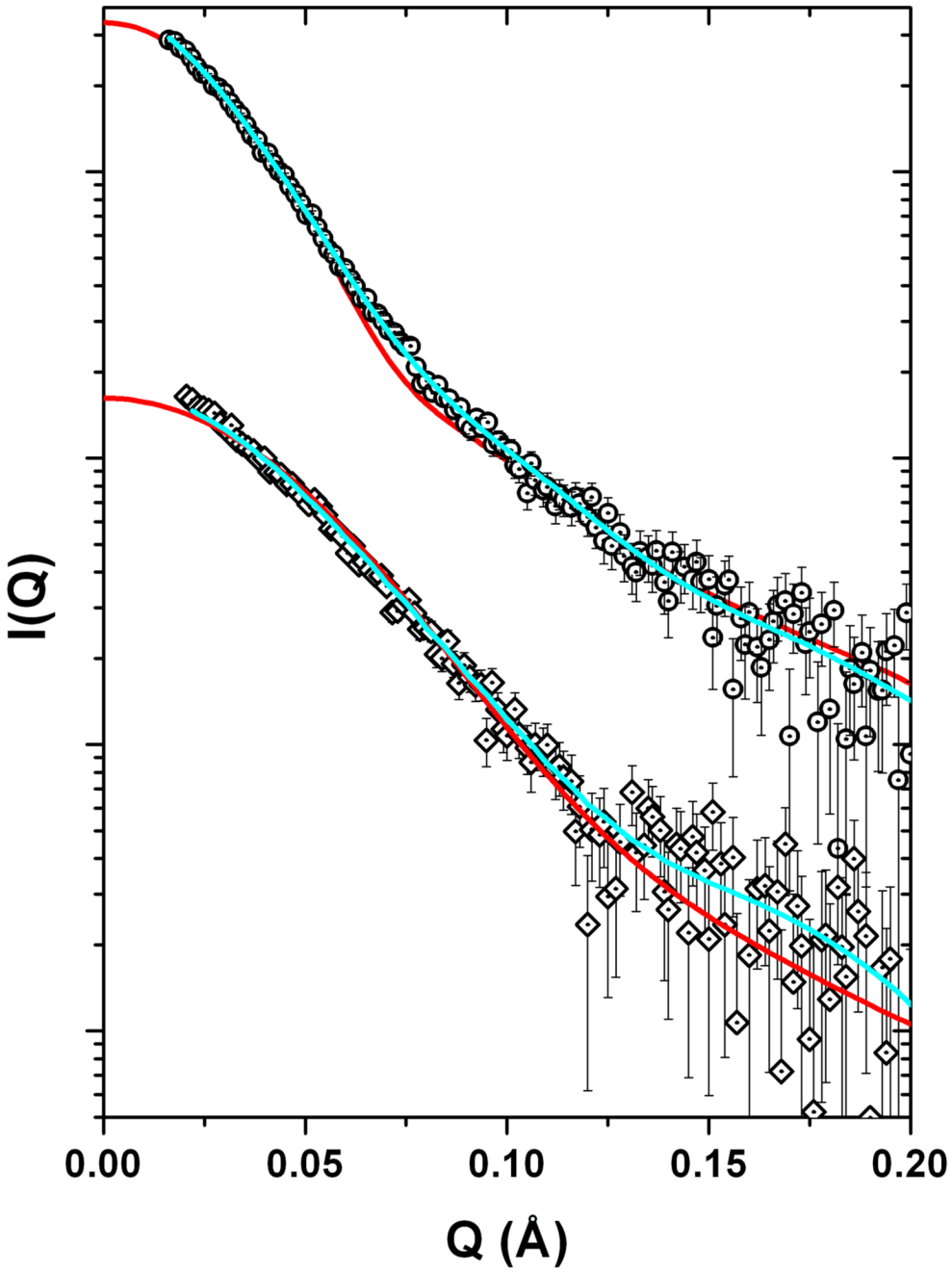


Figure 7. ***I(Q)*** **versus** ***Q*** **for the apo-R3 and native LacI-*o*DNA compared with the crystal structure and best-fit model** ***I(Q)*** **profiles**

The experimental data are the desmeared scattering profiles from the respective data in Figure 2 and the model profiles were calculated using CRYSOL and crystal structure coordinates (1LBG PDB[3]) (red profiles) and the best-fit model profiles (cyan). For the apo-R3 data (lower curves), the BUNCH-derived model illustrated in Figure 6 is the best fit to experiment. For the native LacI tetramer data (upper curves), the best-fit to the experimental data is obtained using an ensemble of structures, illustrated in Figure 8, derived from the crystal structure and accommodating different relative dispositions of the homodimers. The ensemble was generated by rotating the homodimers with respect to each other so as to open the V-shape of the crystal

structure in 20° increments. The model for the ensemble shown is made up of fractional proportions of the crystal structure (0.62) and structures opened progressively by 60° (0.20) and 120° (0.18).

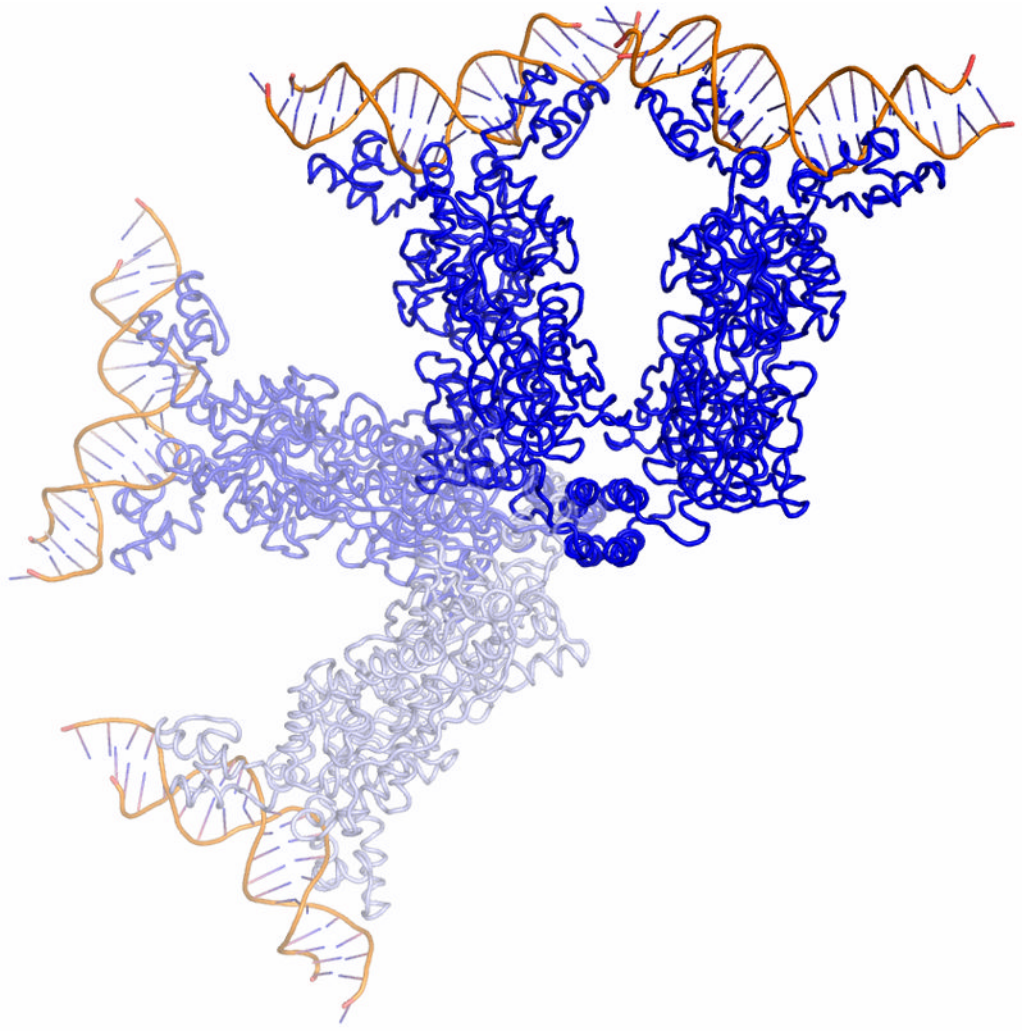


Figure 8. The crystal structure of the native LacI dimer of homodimers with the opened structures superimposed

The dimer subunits of the V-shaped crystal structure (deep-blue ribbons) have been rotated about the tetramerization domain to place dimers at approximately 60 and 120 degrees with respect to the partner dimer (shaded blue and light-blue, respectively, indicating they are increasingly less populated as they become more extended).

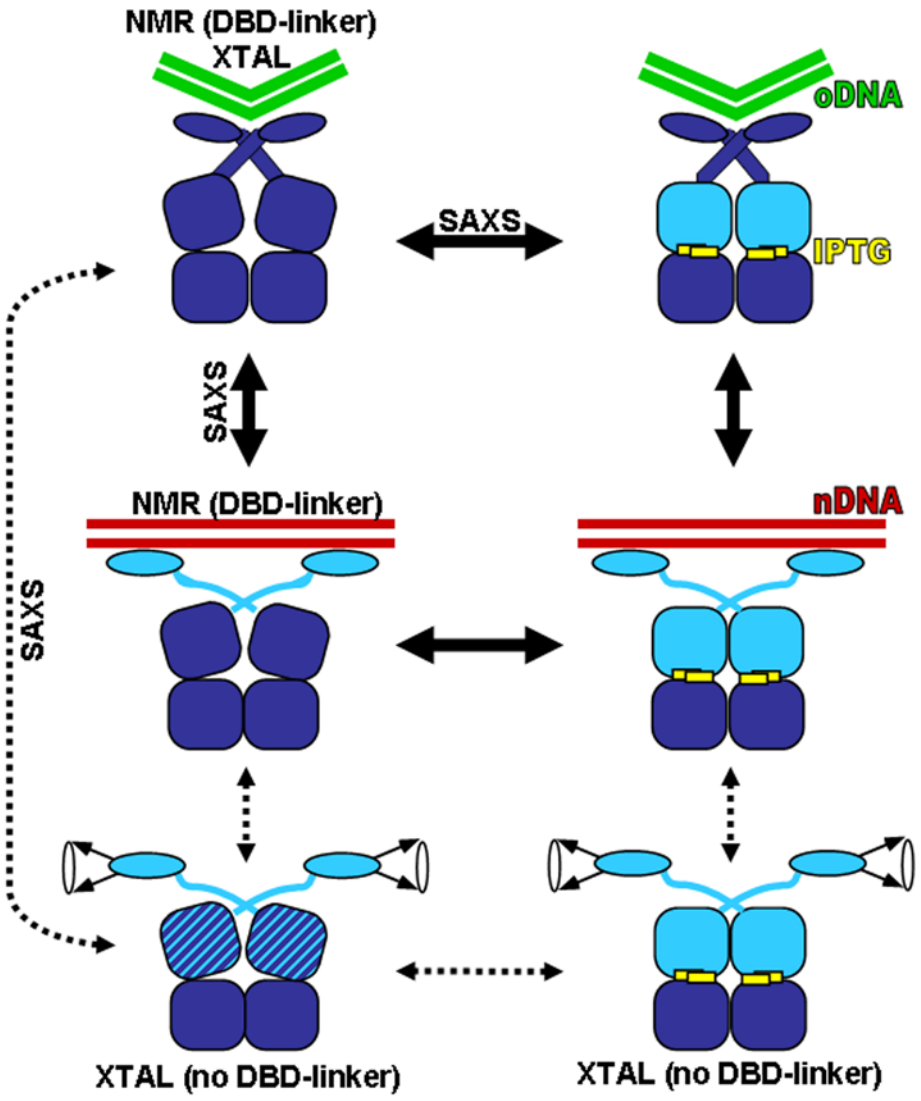


Figure 9. Structural changes during the LacI functional cycle

Probable structures for LacI homodimer complexes that occur during the functional cycle are represented in this cartoon. The method by which experimental evidence for each structure was obtained is indicated on the figure. The LacI homodimer is represented by the blue figures; the DBD are depicted as small ovals at the top, linkers are shown as rectangles, and the two subdomains of the regulatory domain are indicated by squares with rounded corners. Cartoons in the left column correspond to LacI in the absence of IPTG. Cartoons in the right column represent LacI bound to IPTG (yellow squares), which causes a conformational change in the N-subdomain of the regulatory domain that is propagated to the DNA binding domains and lowers affinity for *o*DNA (lighter blue). *o*DNA is represented by the green "V" in the top row. When LacI is bound to *o*DNA, the protein hinge helices are folded and the DNA is kinked [3; 5; 29; 77]. The current solution scattering data indicate that the linker hinge helices remain compact in the ternary LacI-*o*DNA-IPTG complex. *n*DNA is represented by the red lines in the middle row. In this complex, NMR experiments with truncated LacI DBD indicate that the linker hinge helix is unfolded and the DNA is B-form[24]. CD experiments further indicate that the structure of *n*DNA differs from that of *o*DNA when bound to LacI [51]. Solution scattering confirms that the complex with *n*DNA is of lower affinity than the ternary complex with

*o*DNA. The bottom row depicts the structures of LacI in the absence of DNA. The regulatory domains of apo-LacI adopt a third conformation that is different from either the DNA- or IPTG-bound forms [3; 78]. Solution scattering data indicate that the hinge helices extend and hence must unfold in this complex. Solid arrows indicate that LacI bound to DNA are the predominant forms *in vivo*; dotted arrows indicate the less-populated forms [16].

TABLE 1

Structural Parameters for LacI Tetramer and Dimer Forms
Part A. Experimentally derived parameters
Part B. Model derived parameters[g]

Sample[a]	Protein conc. (M×10^4)	R_g (Å)	d_{max}[b] (Å)	V[c] (Å^3)	V_{calc}[d] (Å^3)	I(0)$_{norm}$[e] I(0)$_{Lys}$	Fit[f]
LacI Homodimer							
Apo-R3	0.32	35.8±0.5	110	106082±9078	99393	1.02±0.19	0.96
R3-*o*DNA	0.72	34.6±0.3	110	84576±4042	112002	0.86±0.15	0.96
R3-*o*DNA-IPTG	0.80	35.6±0.2	110	88204±3333	"	0.88±0.16	0.97
LacI tetramer							
Apo-LacI	0.16	40.4±0.9	135	185253±1471	191980	0.96±0.19	0.95
LacI-*o*DNA	0.21	48,0±0.8	160	207084±9069	217436	0.99±0.19	0.93
LacI-*o*DNA-IPTG	0.26	47.8±07	160	183937±7083	"	1.00±0.18	0.91
LacI-*n*DNA	0.50	46.0±0.5	150	164832±4431	"	0.73±0.14	0.92
LacI-*n*DNA-IPTG	0.50	43.9±0.5	150	162043±4248	"	0.81±0.15	0.90

	R_g (Å)	d_{max} (Å)	χ^2
Homodimer Models			
apo-form: crystal structure	29.80	90	1.8
BUNCH model	34.04	110	1.02
*o*DNA complex: crystal structure.	34.10	110	1.3
Tetramer Models			
apo-form: crystal structure	38.2	120	1.2
*o*DNA complex: crystal structure	43.4	135	5.23
SASREF model	44.8	145	1.96
ensemble model	47.8	150	0.80

[a] Stoichiometry for Protein/DNA/effector samples is always 1 duplex DNA per LacI homodimer, and 1 IPTG per LacI monomer; i.e. native LacI binds 2 duplex DNAs and 4 IPTG molecules while R3 binds 1 duplex DNA and 2 IPTG. *o*DNA refers to operator duplex DNA, *n*DNA to nonspecific duplex DNA. Both *o*DNA and *n*DNA are 21 base pairs.

[b] The d_{max} values are a good approximation to the actual maximum linear dimension for the all protein samples, but for the complexes with DNA the more electron dense DNA at one end of the particle shortens the d_{max} value calculated from the scattering data compared to the actual maximum dimension of the particle.

[c] Calculated using the Porod invariant for globular particles [40].

[d] The volumes calculated from the protein and DNA sequence based on Voronoi volumes [41].

[e] $I(0)_{norm}/I(0)_{Lys}$ is the ratio of scattering intensities of the sample and a lysozyme standard normalized by respective concentrations (mg/ml), molecular weights, and scattering contrast measured in the same capillary and should be 1.00 for monodisperse particles. Errors are based on propagation of counting statistics plus contributions from relative errors in the determination of protein concentration (see Materials and Methods).

[f] Total quality estimate defined in GNOM, equal to 1.0 for an ideal solution [76].

[g] Parameters for models are from CRYSOL [74] using crystal coordinates derived from the LacI tetramer, PDB 1LBG [3]. Homodimer structures used only chains AB (apo) and ABEF (DNA complex). For R_g the standard hydration layer estimates were used. The d_{max} values were estimated using the model profiles and GNOM to calculate model $P(r)$ profiles in a similar manner to the experimental data. The χ^2 values quoted are for the fits of the CRYSOL model intensity profiles to the appropriate desmeared experimental scattering profiles from the data shown in Figure 2.

TABLE 2

Guinier Analysis for Rod-Shaped Particles for LacI and its variants

Sample	*C* (mg/mL)	R_g (Å)	R_c*a* (Å)	$R_{c(Xtal)}$*b* (Å)	*I(0)/c*
Apo-R3	21.0	56.3±1.8	16.8±0.2		99.3±11.5
	15.5	54.6±3,7	16.6±0.3		97.0±11.5
	8.9	52.2±1.4	16.4±0.2		83.7±10.0
	3.7	44.1±2.6	16.7±0.3		71.1±8.5
	2.6	32.6±0.9	16.8±0.3	16.7	104.7±12.4
Apo-LacI	28.1	66.6±0.9	23.3±0.6		152.5±25.9
	18.8	59.8±0.8	23.2±0.6		154.1±23.0
	13.8	54.8±0.7	23.8±0.3		164.9±21.0
	8.4	51.2±0.8	23.8±0.4		181.8±24.0
	2.5	38.9±0.7	23.4±0.5		161.7±21.7
	0.9	38.6±1.1	23.5±0.9	26.0	167.4±27.0

[a] R_g and R_C values were obtained from the scattering data using GNOM[42].

[b] $R_{C(Xtal)}$ values are calculated in the same manner as for the experimental data, but using model scattering profiles based on crystal structure coordinates from the PDB structure 1LBG and using the program CRYSOL[74]. For apo-R3 comparison coordinates for one protein homodimer were used and for apo-LacI the tetrameric protein coordinates were used.

[c] Errors are propagated statistical errors plus estimated 10% error in protein concentration.

TABLE 3

Model fits to the Native LacI Dimer of Dimers complexed with *o*DNA

Component structures included in the population ensemble[a]	χ_2 (fit of model profile to experiment)	R_gb (Å)	Populations of component structures in th
Crystal structure (1LBG)	5.23	43.4	1.0
Crystal + 20° open	3.25	45.6	0.53 ± .05, 0.47 ± .05
Crystal + 40° open	1.56	46.9	0.60 ± .02, 0.40 ± .02
Crystal + 60° open	0.89	47.6	0.62 ± .01, 0.38 ± .01
Crystal + 80° open	0.82	47.9	0.63 ± .01, 0.37 ± .01
Crystal + 100° open	0.87	47.9	0.64 ± .01, 0.36 ± .01
Crystal + 120° open	0.91	47.8	0.64 ± .01, 0.36 ± .01
Crystal + 140° open	0.90	47.8	0.63 ± .01, 0.37 ± .01
Crystal + 160° open	0.86	47.8	0.62 ± .01, 0.38 ± .01
Crystal + 180° open	0.80	47.8	0.59 ± .01, 0.41 ± .01
Crystal + 60°, 120°	0.80	47.8	0.62 ± .01, 0.20 ± .04, 0.18 ± .04
Crystal + 60°, 160°	0.82	47.9	0.63 ± .01, 0.28 ± .11, 0.09 ± .11
Crystal + 80°, 180°	0.81	47.8	0.60 ± .23, 0.08 ± .16, 0.32 ± .18

[a] Populations with one of the more open structures all gave χ^2 values > 6. With more than 3 component structures excellent χ^2 values can be obtained, but generally with negative coefficients indicating unphysical population distributions.

[b] Estimated from Guinier analysis of the model profiles with QR_g<1.3.

TABLE 4

Thermodynamic characterization of LacI and R3 DNA binding.

LacI Variant	Buffer	K_d: *o*DNA (M×10^{11})	[IPTG]$_{mid}$[a] (M×10^6)
Wild-type LacI	FBB[b]	1.5 ± 0.4	
	SAXS[c]	25 ± 2.8	1.5 ± 0.2
R3	FBB	1.3 ± 0.3[d]	
	SAXS	58 ± 14	2.5 ± 0.3

[a] $[IPTG]_{mid}$ reflects the concentration of IPTG required to release 50% of *o*DNA from the high affinity complex.[68]

[b] FBB buffer is: 0.01 M Tris, pH 7.4, 0.15 M KCl, 0.3 mM DTT, 0.1 mM EDTA, 5% DMSO. Binding is to 40-mer *lacO1*, which corresponds to the current *o*DNA with ~10 extra basepairs on either side of the binding site. Average and reported error (standard deviation) are determined from at least three independent determinations, as reported in reference [67].

[c] Affinity was determined in the same buffer as the solution scattering experiments: 100 mM Tris-HCl, pH 7.4, 100 mM KCl, 1 mM EDTA, and 0.3 mM DTT. Affinity was determined using *lacO1* 40-mer. Solution scattering was carried out with *lacO1* 21-mer (*o*DNA). The reported values are the average of three determinations; reported error is the standard deviation.

[d] Data newly acquired for the current R3 protein. Value is in good agreement with published value of $(3.3 \pm 0.3)\times10^{-11}$ M. [32] Reported error is error of the fit for an individual determination.