Using Networks To Identify Fine Structural Differences between Functionally Distinct Protein States†

Liskin Swint-Kruse*

Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, MS 3030, Kansas City, Kansas 66160, and Department of Biochemistry and Cell Biology, W. M. Keck Center for Computational Biology, MS 140, Rice University, Houston, Texas 77005

Received March 19, 2004; Revised Manuscript Received May 13, 2004

ABSTRACT: The vast increase in available data from the "-omics" revolution has enabled the fields of structural proteomics and structure prediction to make great progress in assigning realistic three-dimensional structures to each protein molecule. The challenge now lies in determining the fine structural details that endow unique functions to sequences that assume a common fold. Similar problems are encountered in understanding how distinct conformations contribute to different phases of a single protein's dynamic function. However, efforts are hampered by the complexity of these large, three-dimensional molecules. To overcome this limitation, structural data have been recast as two-dimensional networks. This analysis greatly reduces visual complexity but retains information about individual residues. Such diagrams are very useful for comparing multiple structures, including (1) homologous proteins, (2) time points throughout a dynamics simulation, and (3) functionally different conformations of a given protein. Enhanced structural examination results in new functional hypotheses to test experimentally. Here, network representations were key to discerning a difference between unliganded and inducer-bound lactose repressor protein (LacI), which were previously presumed to be identical structures. Further, the interface of unliganded LacI was surprisingly similar to that of the K84L variant and various structures generated by molecular dynamics simulations. Apo-LacI appears to be poised to adopt the conformation of either the DNA- or inducer-bound structures, and the K84L mutation appears to freeze the structure partway through the conformational transition. Additional examination of the effector binding pocket results in specific hypotheses about how inducer, anti-inducer, and neutral sugars exert their effects on repressor function.

The fields of protein structure prediction and structural proteomics are making great strides toward assigning each primary sequence a satisfactory three-dimensional structure. However, as these disciplines advance toward their goal, other studies appear with increasing frequency that illustrate the next challenge in understanding the protein structure/function relationship (e.g., refs *12*−*14*): discerning structural details that impart unique function to each of these structures. These fine details not only influence the varied functions of homologous proteins but are also essential for comprehending how functionally distinct states of the same protein differ in their conformations. Since protein structures are complicated, identifying small but functionally relevant details can be extremely difficult. A common approach couples C^{α} alignments with visual inspection and rmsd[1] calculations (e.g., Figure 1). However, the work presented in this paper illustrates how functionally important details can be missed with this approach. Overlooked detail is not limited to altered side chain conformations: one of the examples reported herein involves a backbone change for about three residues in a 300 amino acid monomer structure of the core domain of the lactose repressor protein (LacI). This detail was uncovered by representing three-dimensional structural data as two-dimensional networks.

LacI has long served as the prototype for protein control of gene regulation (*1*). Throughout the 1960s, 1970s, and 1980s, genetic and biochemical approaches greatly contributed to understanding the function of this protein (e.g., refs *1* and *15*−*30*).[2] However, the structure of LacI proved elusive. Some features could be inferred using biochemical labeling, domain truncation, amino acid substitution, and homology modeling (e.g., refs *5*, *6*, *22*, *24*, and *31*−*34*), but interpretation of the results was limited by the lack of high-resolution structures. Finally, after decades of frustration, crystallographic and NMR structures were determined for a variety of LacI liganded states (Table 1) (*8*−*10*, *35*−*40*).[3]

† This work was supported by NIH Grant GM22441 and Robert A. Welch Grant C-576 to Kathleen Shive Matthews (Rice University). L.S.-K. was supported in part by a fellowship from the W. M. Keck Center for Computational Biology (National Library of Medicine Grant LM07093).

* To whom correspondence should be addressed. Telephone: (913) 588-7008. Fax: (913) 588-7440. E-mail: lswint-kruse@kumc.edu.

[1] Abbreviations: rmsd, root mean square distance; LacI, lactose repressor protein; IPTG, isopropyl β-D-thiogalactoside; ONPF, *o*-nitrophenyl β-D-fucoside; TMD, targeted molecular dynamics; CSU, Contacts of Structural Units; LPC, Ligand Protein Contacts; CE, Combinatorial Extension.

[2] When the LacI protein binds to a specific operator DNA sequence, transcription of downstream genes is repressed. "Induction" of gene expression occurs when LacI binds a small "inducing" sugar molecule at a binding site distant from the DNA binding site. Even though the sites are far apart in space and primary sequence, the information of inducer binding is transmitted through the protein and results in allosterically lowered operator affinity (>1000-fold). For review, see Matthews and Nichols (*1*).

10.1021/bi049450k CCC: $27.50
Published on Web 08/04/2004

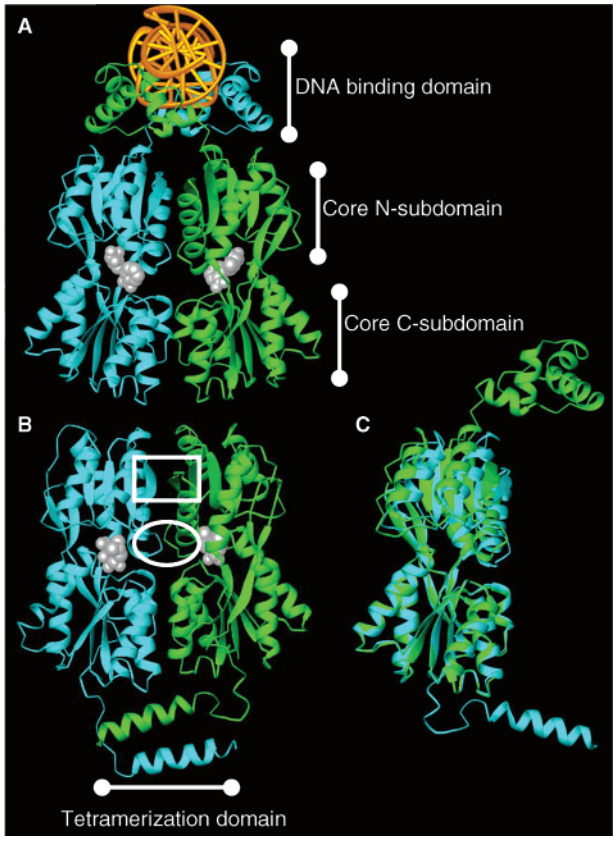

FIGURE 1: Structures for LacI bound to (A) O^{sym} DNA (gold) and anti-inducer ONPF (white balls) (dimer variant with no tetramerization domain; PDB ID 1efa) (*8*) and (B) inducer IPTG (white balls) (one dimer of a tetramer; no electron density is detectable for the N-terminal DNA binding domains; PDB ID 1lbh) (*40*). The white oval indicates the position of the 70's region, while the white rectangle marks the 80−100's region. In (A) and (B), individual monomers of a dimer are colored green and cyan. (C) Monomers from 1efa (green) and 1lbh (cyan) were aligned using residues in the core C-subdomains. Note the movement of the N-subdomains relative to one another.

Although the intact repressor was initially crystallized in three functional states (DNA bound, inducer IPTG bound, and unliganded), attention turned immediately to the differences between the liganded conformations (Figure 1) (*8*, *40*−*42*). The structure of the unliganded protein appeared to be identical to IPTG-bound LacI (*40*), and thus the apo repressor structure was largely ignored. However, the close similarity between the apo and induced conformations was troubling, since both spectroscopy and chemical modification experiments detected a conformational change when unliganded LacI binds inducer (e.g., refs *22*, *24*, and *43*−*45*). Nonetheless, an easy explanation for the discrepancy between structure and experiment was that the process of crystallization trapped a conformation that is thermodynamically accessible in solution but usually populated at low frequency. This problem was therefore not explored further.

A LacI structural change is required to explain the allosterically altered ligand binding observed for the repressor functional cycle. The complete transition is not yet clear, since all LacI forms crystallized in the absence of DNA do

[3] Crystallographic structures verified the tetrameric nature of LacI known from biochemical studies (reviewed in ref *1*). However, tetramerization is primarily accomplished through a C-terminal four-helix bundle (*2*−*5*), and thus this structure can be described as a "dimer of dimers". In fact, the dimeric unit has the same intrinsic DNA and inducer binding properties, including both allosteric response between the two separate ligand binding sites and cooperativity between two inducer binding sites of a dimer (*6*, *7*). The dimer variant of LacI has proved easier to crystallize (*8*−*10*), and this version is now frequently utilized for structural studies.

Table 1: Available Crystal Structures That Include the Core Domains of LacI

PDB ID, resolution (Å)	description	rmsd with 1lbh[a]
1tlf[b]	tetrameric core domains bound to inducer IPTG	1.02
2.60	generated by proteolytic removal of the DNA binding domains	
1lbh[c]	tetrameric core domains bound to inducer IPTG	0.41
3.20	no electron density for mobile DNA binding domains	
1lbi[c]	tetrameric core domains of unliganded protein	0.70
2.70	no electron density for mobile DNA binding domains	
1lbg[c]	tetramer bound to symmetric operator DNA (O^{sym})	
4.80	C^{α} carbons only, no side chains	
1efa[d]	bound to symmetric operator O^{sym} and anti-inducer ONPF	2.85
2.60	dimeric variant, tetramerization domain deleted	
1jwl[e]	bound to natural operator O^1 and anti-inducer ONPF	3.23
4.00	dimeric variant, tetramerization domain deleted	
1jyf[f]	glycerol in inducer binding pocket	1.06
3.00	dimeric variant (no tetramerization domain); no electron density for mobile DNA binding domains	
1jye[f]	glycerol in inducer binding pocket	2.37
1.70	K84L mutation/dimeric variant (no tetramerization domain); no electron density for mobile DNA binding domains	

[a] Values for rmsd of the N-subdomain were determined after aligning the core C-subdomain structure of monomer A with the same region of monomer A from the structure 1lbh. The alignment was accomplished with the Web-based program CE (*53*). The rmsd value for 1lbh was calculated by comparing A and B monomers of this tetrameric structure. [b] Reference *35*. 1tlf: $P2_1$ spacegroup; *R*-value 0.222; four monomers in the asymmetric unit; temperature factors range from 4 to 150 over the entire structure, whereas those in the N-subdomain interface and the inducer binding pocket are generally between 10 and 50. [c] Reference *40*. 1lbh: $P2_122_1$ spacegroup; *R*-value 0.230; four monomers in the asymmetric unit; temperature factors for the entire structure range from 2 to 34; those in the N-subdomain interface generally fall in the range of 2−14, and those of the inducer binding pocket are <10. 1lbi: *C*2 spacegroup; *R*-value 0.250; four monomers in the asymmetric unit; temperature factors for this structure are not reported (values in the appropriate column of the PDB file are all identical). 1lbg: *C*2 spacegroup; *R*-value 0.260; four monomers in the asymmetric unit, C^{α} atoms only. [d] Reference *8*. 1efa: *H*32 spacegroup; *R*-value 0.247; three monomers in the asymmetric unit; temperature factors for the entire molecule range from 30 to 130, whereas those in the N-subdomain interface are between 40 and 70, and those of the inducer binding pocket fall in the range of 40−55. [e] Reference *9*. 1jwl: *H*32 spacegroup; *R*-value 0.249; three monomers in the asymmetric unit. Temperature factors for the N-subdomain interface range from 30 to 70 while those of the DNA binding domain (where some side chains cannot be determined) are generally >130. Temperature factors for the inducer binding pocket are generally around 40. [f] Reference *10*. 1jyf: $I4_122$ spacegroup; *R*-value 0.203; one monomer in the asymmetric unit; temperature factors range from 20 to 100 across the structure, whereas those in the N-subdomain interface are generally between 40 and 60, and those of the binding pocket are between 20 and 45. 1jye: $I4_122$; *R*-value 0.223; one monomer in the asymmetric unit; temperature factors range from 11 to 61 across the monomer, whereas those in the N-subdomain interface are between 18 and 30, and those of the binding pocket are usually <20.

not have electron density for the mobile DNA binding domains. Nonetheless, the sugar binding core domains evince a structural shift that is capable of disrupting the orientation of the DNA binding domains, precluding protein−protein contacts requisite for high-affinity DNA binding (Figure 1C) (*8*, *40*, *46*). The core domain can be divided into two interconnected N- and C-subdomains. During the shift, the C-subdomains are stationary and anchor the motions of the N-subdomains (*8*, *40*). All studies that compared the DNA- and inducer-bound conformations identified the N-subdomain orientation and interface as a critical difference between the two conformations.

One prominent participant in the N-subdomain interface rearrangement is the side chain of Lys 84 (*10*, *42*, *47*, *48*). A structure has been determined for the K84L dimer variant of LacI and compared to the dimeric version of wild-type protein, with previous interpretation based on the assumption that the structures were in the uninduced state (*10*). On recent reexamination of these structures, a glycerol molecule (included in solution for cryoprotection) was noted to be in the inducer binding pocket in a manner very similar to inducer IPTG. Although present in the PDB coordinates, this feature had been dismissed since glycerol was classified as a "neutral" sugar (*29*), serving as neither inducer nor anti-inducer.[4] A forthcoming paper presents the results of new biochemical experiments that verify specific glycerol binding to LacI in the absence of induction (or anti-induction) as well as impaired inducibility for LacI variants at position 84 (Swint-Kruse, Zhan, and Matthews, unpublished results). This paper reports previously unnoticed structural distinctions that correlate with functional features of the eight available LacI experimental structures (including that for K84L). These studies also make predictions about LacI conformations that are not yet known with high resolution.

A key tool to finding these differences is the two-dimensional network representation of the N-subdomain interface (Figure 2). Similar analyses were previously helpful in two other situations where multiple structures were compared: (1) exploring differences between the interfaces of homologous proteins (*41*) and (2) following conformational change during a dynamics simulation (*50*). This approach is similar to the three-dimensional network analysis of protein structure recently reported by Greene and Higman (*51*). Network representation retains information about individual amino acids yet simplifies the visual complexity of three-dimensional representations, which impairs direct comparison of more than two or three structures. This graphical display allows facile, simultaneous comparison of multiple structures which can direct the investigator toward interesting regions for detailed structural examination and future experiments.

METHODS

The two-dimensional networks were generated as described in Results and Discussion as well as in Swint-Kruse et al. (*50*). The work relied heavily upon the Web-based programs Contacts of Structural Units (CSU) and Ligand Protein Contacts (LPC) (http://bip.weizmann.ac.il/oca-bin/lpccsu) (*52*). Visual inspection utilized Rasmol (http://www.umass.edu/microbio/rasmol/index2.htm). Protein alignments were accomplished with the Web-based program Combinatorial Extension (http://cl.sdsc.edu/ce.html) (*53*). Protein structures are represented by the program Ribbons (http://sgce.cbse.uab.edu/ribbons/) (*54*). A strategy for using these tools is described in Swint-Kruse and Matthews (*55*). Briefly, Rasmol was utilized to generate a list of all atoms on one monomer within 4.5 Å of the partner subunit. Residues indicated were explored more fully with CSU, to determine the positions of their interacting partners and to narrow down those interactions that meet the distance criteria (any two atoms within 3.5 Å or hydrophobic side chain atoms within 4.5 Å). These data were then incorporated into the networks. The "equilibrated" structure generated from PDB ID 1lbh (*40*) in the absence of IPTG was originally reported in ref *42*. This process was executed using CHARMM (*56*), and the extended equilibration reported in this paper simply increased the final molecular dynamics step from 10 to 100 ps.

RESULTS AND DISCUSSION

Strategies for Comparing Protein Structures. As the number of available protein structures expands exponentially, detailed examination of all related structures becomes a daunting task. Even with the availability of Web-based programs for structure-based alignments, such as Combinatorial Extension (*53*), visual inspection of the similarities and differences is impaired by the size and complexity of protein molecules. Illustrations with even three aligned protein backbone ribbons with all atoms shown for four side chains can obscure relevant detail. If the proteins of interest have multiple domains or interconnected subdomains, choosing a subset of residues for the basis of the alignment further complicates analysis (e.g., ref *50*).

Alignment algorithms do calculate a C^{α} root mean square distance (rmsd) value, which essentially expresses the degree of similarity between two structures as a single number. However, for even moderately large proteins, this number can obscure the fine structural details that might be critically important to functional differences (*57*). Further, even if domain reorientation is incorporated into the calculation, critically important differences may be missed. For example, Table 1 presents rmsd values calculated for the N-subdomains of LacI after aligning the C-subdomains using IPTG-bound 1lbh as a reference. Note that this value for 1lbi (apo-LacI) is lower than even for the other two structures bound to IPTG (1lbh second monomer and 1tlf), masking a critical structural difference in the N-subdomain between the two conformations (see below).

Two-dimensional networks are very useful for comparing a large number of structures to each other, such as members of a homologous family (*41*) or representative time points from a conformational change (*50*) (e.g., Figure 2; previously called "contact maps" in ref *41*). This approach to structure comparison provides a compromise between two extremes, reducing the visual complexity of a protein model but maintaining separate information (including data for the side chain) about each residue. In constructing the network, participating residues from each monomer are indicated by their sequence numbers and arranged in parallel columns.

[4] Note that anti-inducing sugars allosterically *enhance* DNA binding (*11*), and the anti-inducer ONPF is in fact present in the high-resolution DNA-bound structures of LacI (*8*, *9*).

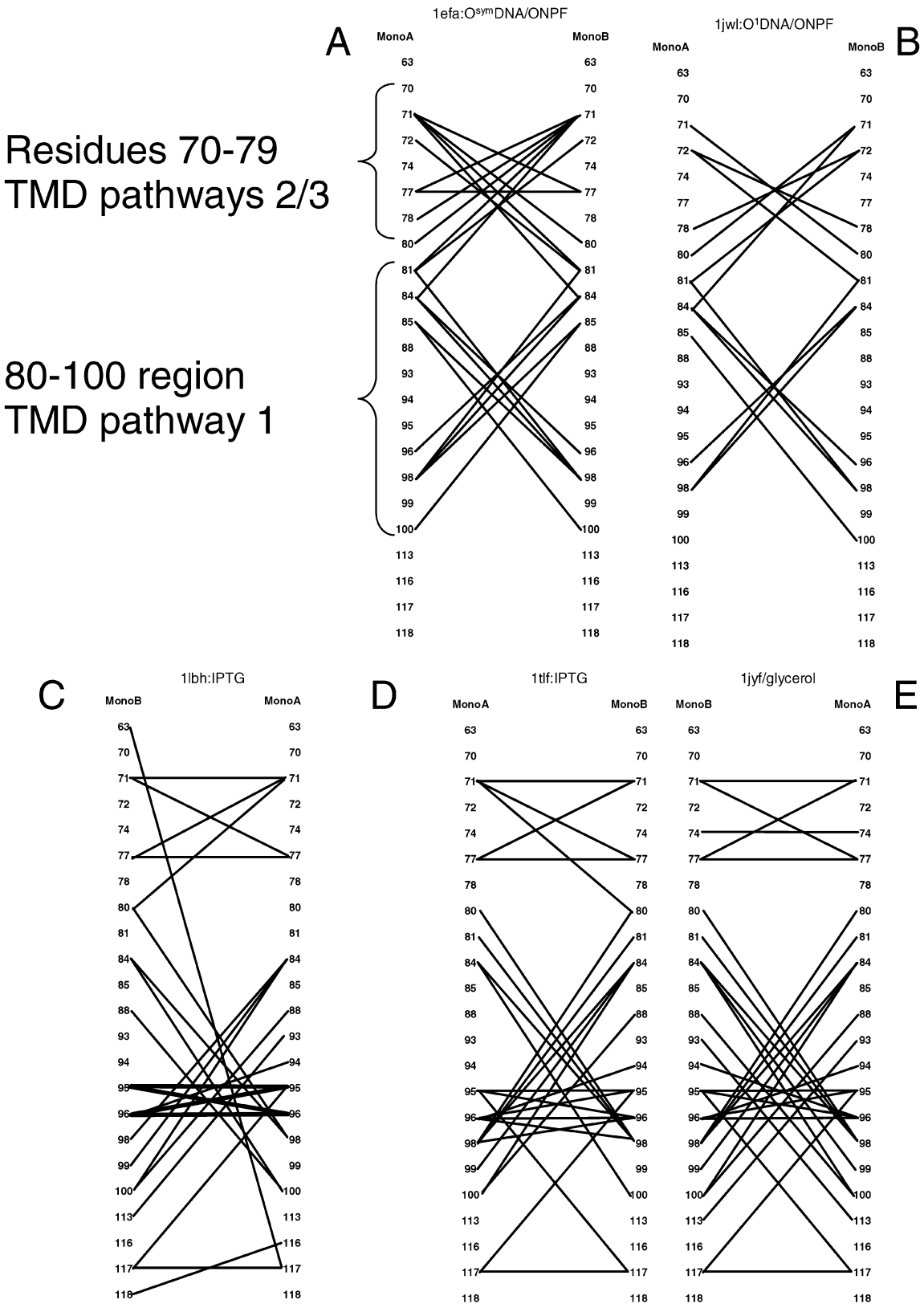

FIGURE 2: Interaction networks for DNA/ONPF-bound (A, B) and IPTG- or glycerol-bound (C−E) wild-type LacI. PDB ID codes and bound ligands are indicated at the top of each network. Each column of numbers represents amino acids that participate in the N-subdomain interface for each monomer. Lines indicate cross-monomer interactions. On (A), the regions involved in the TMD pathways are noted to the left of the network (*42*). The heavy dark lines on the network for 1lbh (B) highlight interactions that form a continuous β-sheet across the N-subdomain interface (*35*, *40*); these interactions are absent in the DNA/ONPF-bound structures.

Interactions between residues of the different monomers are indicated by lines. For this purpose, an "interaction" is defined by a rather broad description: if any atoms of the two residues are $\leq$3.5 Å apart, or if hydrophobic regions of side chains are $\leq$4.5 Å, an interaction is indicated. Additional information can be noted by changing the style of the

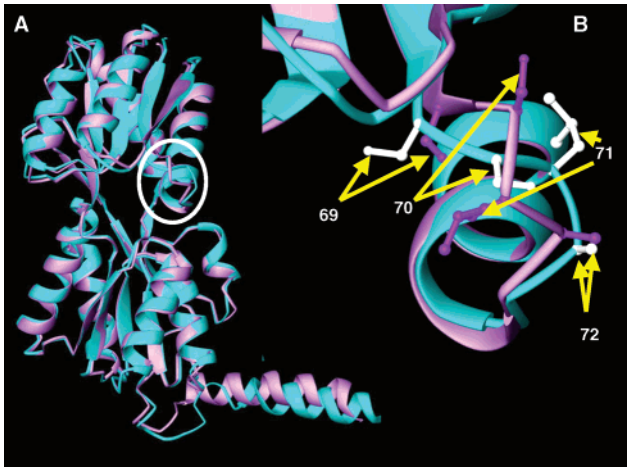

FIGURE 3: Comparison of IPTG-bound 1lbh (cyan, white sticks) and unliganded 1lbi (lavender, purple sticks) (*40*). (A) Structures were aligned using only the residues of the C-subdomain. Note the high similarities of the N-subdomains. The white circle indicates the area enlarged in panel B. (B) Discrepancies in the backbones and side chains for N-subdomain residues 69−72.

connecting line (e.g., Figure 2C, heavy black lines). This representation should also be amenable to "network analysis", similar to the study of three-dimensional networks recently generated to explore the types of contacts in protein structure (e.g., long range vs short range; small world vs random) (*51*).

One caveat that must be kept in mind is that protein structures are simply models built from crystallographic data and thus contain inherent experimental and modeling error. Indeed, most of the LacI structures have moderate resolution (Table 1). However, the interface differences noted below are due to large rearrangements of complete side chains (Figure 3B). These changes were missed in traditional analyses because (1) in an all-atom representation of these structures, side chain movements for 3 out of 300 residues are not obvious if the investigator does not have reason to focus on this area and (2) the large number of C^{α} atoms incorporated into rmsd calculations obscures the changes that occur in the backbones of these sites. For the binding pocket differences, entire side chains also make significant movements (Figure 5), and the "interaction distance" (the shortest point between a side chain and the ligand) changes from 1 to ≥2.5 Å for binding ONPF compared to other ligands (Table 2, dotted). These differences are greater than the uncertainty of the models, which are estimated by Luzzati diagrams to be greater than 0.35 Å but less than 0.5 Å for these models (see Table 1 for resolution and *R*-values) (*49*). Further, temperature factors for the 70−80 region of the N-subdomain interface (where the most profound differences are noted) are some of the lowest in the structures (Table 1, footnotes).

Importantly, all of the structural details highlighted in this paper are noted multiple times within the presented data. Features of the binding pocket are replicated in two different structures for each ligand bound. DNA- and inducer-bound interfaces have been experimentally determined at least twice in different resolution structures, and the interface *pattern* observed for apo-LacI recurs three times: (1) in the K84L variant, (2) at the end of a molecular dynamics trajectory simulating the conformational change that accompanies IPTG dissociation from the inducer-bound LacI, and (3) at intermediate time points in the TMD trajectory between the DNA- and inducer-bound states. Most intriguingly, the comparisons of this paper predict a structural conformation for the region connecting the inducer binding site and the N-subdomain interface (amino acids ~69−75) when LacI is bound to DNA but *not* anti-inducer. Indeed, this hypothesis is consistent with the low-resolution structure of the same (see Conclusions). Of course, as with all structural observations, predictions should be tested experimentally.

Table 2: Distances from LacI Contact Residues to the Sugar Ligand[a]

Residue number	S69	L73	H74[d,e]	A75	P76	I79	R101	N125
Phenotype[b]	I^{ws}	I^{s}	~I^{s}	I^{s}	I^{-}/I^{ws}	I^{-}/I^{ws}	I^{+}	I^{s}
1EFA:A/ONPF[c]		3.9	4.4	3.1	3.9	3.2		3.6
1JWL:A/ONPF		4.1	4.5	3.1	4.1	3.1		4.2
1LBH:A/IPTG	3.3	4.7		3.6	4.1	3.4	5.1	4.4
1TLF:A/IPTG	4	3.9		3.2	3.3	3.9		4
1JYF/glycerol		5.3		3.7	5.5			
1JYE/glycerol			5.5	3.9	5.8			

Residue number[a]	P127	L148[f]	D149[e]	F161	S191[e]	S193	R197[g]	W220[h]
Phenotype[b]	I^{s}	~I^{-}	I^{s}	I^{-}/I^{ws}	I^{s}	I^{s}	I^{s}	I^{s}
1EFA:A/ONPF		3.3	3.1	3.4	4.2	3.6	3.2	4
1JWL:A/ONPF		3.4	3.6	3.3	4.9	4.1	3.4	3.9
1LBH:A/IPTG	4.9	4.5	3.5	6	3.3	4.4	2.5	3.6
1TLF:A/IPTG	3.9	3.5	3.2	5.3	4.8	3.2	2.8	3.6
1JYF/glycerol			4.8			3.6	3	4
1JYE/glycerol						4.3	2.8	3.8

Residue number[a]	A245	N246[e]	Q248[e]	Y273[i]	D274[e,j]	Q291	F293[i]	L296[e]
Phenotype[b]	I^{-}/I^{s}	I^{s}	I^{s}	I^{s}	I^{s}	~I^{+}	I^{s}	I^{s}
1EFA:A/ONPF		3.1	4.7		2.4	4	3.4	4
1JWL:A/ONPF		3.4	4.5		2.3	3.6	3.4	3.8
1LBH:A/IPTG		3.5	5.3	4.5	2.8	3.5	4.6	5.2
1TLF:A/IPTG		2.8	5		2.6	4.2	4.4	5.4
1JYF/glycerol		3.4			2.5	4	5.6	
1JYE/glycerol	3.8	3.2	4.9		2.6	3.1	4.9	

[a] Shortest distance (in angstroms) between each residue and the indicated ligand was determined by the Web-based program Ligand Protein Contacts (*52*) and agreement with crystallographic papers (*8*−*10*, *35*, *40*). If a distance is not designated, it is >6 Å. [b] Primary phenotype from ref *18*. I^{+} indicates phenotype is unchanged; I^{s} and I^{ws} indicate no or impaired response to inducer; I^{-} denotes lost repression; if preceded by a ~ sign, then only a few substitutions exhibited the phenotype. [c] This column contains the PDB ID for each structure, followed by a designation for the representative monomer presented and the sugar ligand bound. The structures 1jye and 1jyf have only one monomer in the asymmetric unit (*10*). [d] Mutations at this position cause functional shifts to either the repressed or induced states (*61*, *62*). [e] One or more mutations at this position compensate the dimerization defect of the Y282D I^{-} mutation, so that the double mutant is capable of repression (*41*). [f] Mutation at this position appears to cause a functional shift to the induced state (*63*). [g] Mutations at this position diminish inducer binding (*64*). [h] Tyrosine mutation reported in ref *65*. [i] F293W has altered inducer binding and sensitivity to fluorescence quenchers in the presence of inducer (*61*). [j] Mutation of D274 abolishes inducer binding (*66*, *67*).

Overview of LacI Networks. The interface networks for LacI structures 1efa (bound to DNA and anti-inducer ONPF) and 1lbh (bound to inducer IPTG) were originally constructed

to compare homologous proteins (*41*). As expected from the known conformational transformation, the C-subdomain interface shows little change between the two conformations (not shown), whereas the N-subdomain interface is subject to extensive change (Figure 2A,C) (*8*, *40*, *41*, *58*). Results from the targeted molecular dynamics (TMD) simulation of the conformational change between these two structures recently called attention back to the N-subdomain interface, highlighting changes at several residues between positions 70 and 100 (*42*). Interestingly, the N-subdomain networks identify two regions that are easily associated with pathways identified by TMD: Cross-monomer interactions between residues ~70−80 correlate with positions that change during TMD pathways 2/3, and interactions between residues ~80−100 correspond to regions that change at the end of TMD pathway 1 (Figure 2A).

Eight different experimental structures are now available for the LacI core domains (Table 1), and to facilitate their comparison, networks have been created for the additional interfaces. Although the original interfaces were confirmed, one group of structures illustrated a surprising third pattern in their interface networks. This pattern is a "hybrid" of the previously recognized patterns, with interactions in the 70's region similar to those of structures bound to DNA and connections between sequences from 80 to 100 like those of inducer-bound structures. Similar intermediate patterns also occur for structures generated by various computer simulations (*42*). As a group, the experimental and simulated structures indicate that unliganded LacI is poised to quickly accommodate binding either ligand (DNA or sugar) and that the K84L variant "freezes" the repressor in this intermediate state (Swint-Kruse, Zhan, and Matthews, unpublished results).

The Interfaces of DNA/ONPF-Bound LacI. Two structures are available for LacI bound to DNA and anti-inducer ONPF: 1efa and 1jwl (*8*, *9*). Note that while 1jwl has lower overall resolution, the temperature factors in the core N-subdomain are comparable to those of 1efa. The interaction networks created for these structures are similar to each other, although 1jwl has fewer total interactions than does 1efa (Figure 2A vs Figure 2B). A third structure, that of tetramer LacI bound to DNA in the absence of ONPF (1lbg), only reports positions for C^{α} and does not have sufficient resolution to construct an interface network (*40*).

In these networks, the region from 70 to 79 on the N-subdomain interface participates in multiple cross-subunit interactions, with fewer in the 80−100 region than in "induced" structures (Figure 2A,B compared to Figure 2C−E). Residues 81 and 84 are notable in that they contact both regions of the N-subdomain interface of the DNA-bound structures (Figure 2A,B) but only the 80−100's region of the other structures (Figure 2C−E). Both positions have been the subject of mutagenesis studies, and effects of the K84L substitution are discussed later in this paper. The A81V substitution abolishes high-affinity operator DNA binding, slows effector binding kinetics several orders of magnitude, and erases several features of the IPTG UV-difference spectrum (*59*). Thus, the authors postulated that the hydrophobic mutation locks the core domain in an induced conformation. Interestingly, addition of ONPF restores some operator binding (*59*), which is consistent with the mechanism of ONPF action postulated later in this paper.

Interfaces of Inducer-Bound LacI. In contrast, the interface network for 1lbh is characterized by having few interactions between amino acid residues 70−77 and extensive interactions over the range of residues 80−113 (Figure 2C). One feature characteristic in the latter region is the "X" pattern found between residues 95 and 96 (Figure 2C, thick black lines), which reflects the continuous β-sheet that forms across the subunit interface in this conformation (Figure 1B, white rectangle) (*35*, *40*). A structure is also available for the proteolytic core fragment bound to IPTG (1tlf) (*35*), and the interaction network created from this structure is very close to that of 1lbh (Figure 2D). The interface of dimeric LacI bound to glycerol (1jyf) is also very similar to the IPTG-bound structures (Figure 2E). This structure was originally interpreted as if it were unliganded (*10*). Similarity between the glycerol-bound and IPTG-bound structures (Figure 2C−E) and differences with the unliganded LacI structure (Figure 4A; see below) prompted new experiments to verify that glycerol is a neutral sugar (*29*): Although glycerol binds specifically in the LacI pocket, this ligand does not induce the repressor (Swint-Kruse, Zhan, and Matthews, unpublished results). Therefore, the interface pattern alone cannot be used to assign the protein structure to a functional conformation. However, further analysis of the sugar binding pocket yields insight into the different modes of action for ligands that induce, anti-induce (i.e., increase DNA binding affinity), or have no effect (neutral) on the allosteric pathway (see below).

"Intermediate" Interfaces of LacI. Given the extremely high similarities between their structures as determined by C^{α} rmsd (Table 1, Figure 3A) (*40*), the interface network for unliganded LacI (1lbi) was expected to be identical to those of inducer-bound protein. However, this interface was strikingly unique (Figure 4A), with the interactions between the residues from ~70−79 creating a pattern similar to the DNA/ONPF-bound conformations, whereas those between 80 and 100 retained the features of the inducer-bound structures. This observation prompted new comparison of the structures, and indeed the backbones of 1lbh and 1lbi deviate at residues 69, 70, and 71 (Figure 3B). All other regions of the core domain are nearly perfectly coincident, and a reasonable assumption would be that the differences are artifacts of the crystallographic model. However, the backbone deviation is repeated in all four subunits contained in the asymmetric unit of 1lbi. More importantly, the network pattern is highly similar to several others: (1) intermediate time points in the TMD trajectory simulating the structural transition between the two end point conformations (e.g., Figure 4C) (*42*), (2) inducer-bound 1lbh subjected to molecular dynamics calculations in the absence of IPTG coordinates (Figure 4D,E; "equilibrated 1lbh") (*42*), and (3) the experimentally determined structure of the K84L variant bound to glycerol (Figure 4B) (*10*). As a group, these results suggest that unliganded LacI is halfway between 1efa and 1lbh, poised to adopt either conformation when the appropriate ligand is available. This interpretation is consistent with experimental observation that apo-LacI is distinct from IPTG-bound LacI (e.g., refs *22*, *24*, and *43*−*45*) and, thus, satisfactorily resolves the previous discrepancy between structure and experiment.

Both the equilibrated 1lbh and the K84L structures warrant additional discussion. The former is predicated on an observation of the TMD simulation: In preparation for the

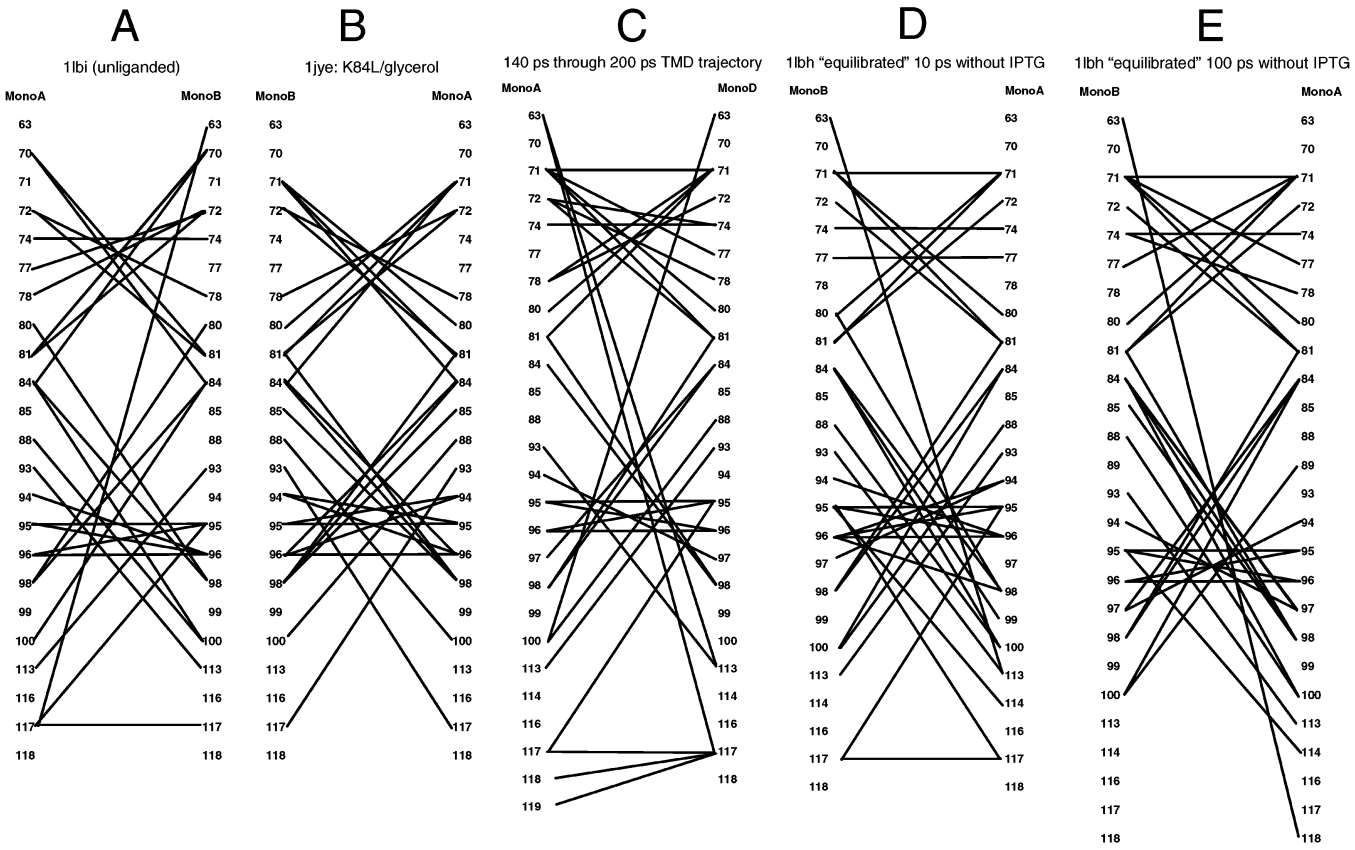

FIGURE 4: Interaction networks for unliganded LacI and the K84L variant bound to glycerol (A, B) and structures generated from various computer simulations (C−E). PDB ID codes and bound ligands are indicated at the top of each network. Each column of numbers represents amino acids that participate in the N-subdomain interface for each monomer. Lines indicate cross-monomer interactions. Note that residues ~70−78 have interactions similar to DNA/ONPF-bound structures (Figure 2A,B). Residues ~80−100 are similar to IPTG-bound structures (Figure 2C−E), including interactions between positions 95 and 96 that are highlighted in Figure 2.

simulation, the inducer-bound structure 1lbh was computationally equilibrated in the absence of inducer IPTG (*42*). Surprisingly, this structure exhibited significant and rapid changes that influenced the subsequent TMD trajectory, and the authors speculated that these changes corresponded to the structure adopting a distinct "apo" state. At the time, most of the differences were noted to occur around residues D149, S193, and R197. However, the interface network for this starting structure also begins to approximate that of unliganded LacI (Figure 4D). To both test the computational hypothesis and confirm the distinct interaction patterns in the region for 70−79 in the experimentally determined, unliganded structure, the molecular dynamics component of "equilibration" was extended from 10 to 100 ps. Gratifyingly, the interaction network becomes significantly more similar to that of unliganded LacI (Figure 4E). Although the C^{α} positions of residues 69−72 are similar to the parent 1lbh instead of the apo 1lbi, the time span of the simulation (100 ps) is at the frequency limit (0.1−10/ns) of such conformational change (*60*).

Both experiment and the TMD trajectory have pointed to the importance of position 84 in the LacI conformational change (*10*, *42*, *47*, *48*). This position has been subjected to mutagenesis, and a forthcoming paper describes additional experiments that demonstrate impaired inducibility of K84A and K84L as well as specific glycerol binding to these mutants (Swint-Kruse, Zhan, and Matthews, unpublished results). Although the relative orientations of the N- and C-subdomains are more similar to the DNA/ONPF-bound conformation (*10*), the interface network for K84L (Figure 4B) is very similar to that of unliganded LacI (Figure 4A). Together, these observations suggest that hydrophobic mutations in this region essentially freeze the structure partway through the conformational change: In terms of the TMD trajectory, pathways 2 and 3 (near and including residues 70−79) can make the requisite changes, but the hydrophobic "bulk" inserted into the interface allows the 80−100 region to interact without a conformational change, thus precluding the motions at the end of pathway 1.

Modes of Action for Inducer, Anti-Inducer, and Neutral Sugars. Studies of the interface networks alone provoke a number of additional questions. First, one limitation of the structures is that the high-resolution DNA-bound structures are also bound to anti-inducer ONPF. Thus, the interface for DNA bound in the absence of sugar ligand may be different from those reported. Second, since their network patterns are similar, one might expect the "intermediate" structures to have the same differences in C^{α} around residues 69−72, analogous to apo-LacI 1lbi. In fact, this region is usually more similar to the IPTG-bound form. Third, why does binding to neutral glycerol make the protein interface look like it is induced when experimentally it is not? By comparing and contrasting the contacts made between repressor protein and the three types of sugar ligands, a number of hypotheses have been formulated about how inducers, anti-inducers, and neutral sugars exert their differential effects on the protein structure.

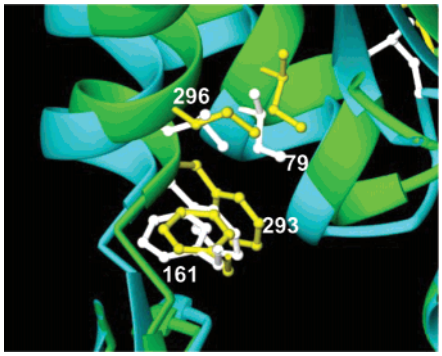

FIGURE 5: Comparison of core pivot side chains for DNA/ONPF-bound 1efa (green, yellow sticks) (*8*) and IPTG-bound 1lbh (cyan, white sticks) (*40*). The side chains of F161, F293, and L296 interact with anti-inducer ONPF, forming a hydrophobic wedge that props the N-subdomain in the open conformation with high affinity for operator DNA.

Distances for protein−sugar contacts are presented in Table 2. An interesting note is that 1tlf and 1lbh model IPTG in the pocket in different orientations (*35*, *40*), but both models utilize very similar protein contacts to inducer (Table 2). Seven repressor positions have very similar interaction distances and are clearly utilized for specific binding of all three sugar ligands: A75, S193, R197, W220, N246, D274, and Q291 (Table 2, light gray). With six structures to survey, a mode of action for anti-inducer ONPF also becomes very clear (Table 2, dotted): The shortest distance between the F161 side chain and the ligand is 2 Å or more longer for IPTG/glycerol than ONPF. For F293 and L296, the shortest contact to anti-inducer is at least 1 Å shorter than to other ligands. Structure examination shows that the entire hydrophobic side chains for residues F161, F293, and L296 are in much closer contact to ONPF than to either IPTG or glycerol. These side chains are actually pulled deeper into the sugar binding pocket in the presence of ONPF (Figure 5) and appear to "prop open" the N-subdomain, thus stabilizing a conformation with high DNA affinity.

Discerning contacts that communicate "induction" is not as obvious, but several points may be deduced. The region of LacI comprising residues 69−79 is intimately involved in both sugar binding and the N-subdomain interface: Residues S69, L73, A75, P76, and I79 contact ligand, whereas S70, L71, A72, H74, S77, and Q78 participate in the cross-monomer interface. Further, only IPTG contacts S69 (Table 2, black); neither anti-inducer ONPF nor neutral glycerol contacts this residue. Given that the backbones of residues 69−72 are strikingly different between inducer-bound and unliganded LacI, an original expectation was that this would be true for all uninduced structures. Thus, the fact that only 1lbi has a different backbone conformation was surprising (with the exception of the low-resolution DNA-bound structure 1lbg; see Conclusions below). However, differential ligand interactions with S69 remain a striking difference between the induced and uninduced structures. The DNA binding domain comprises residues 1−62, with L62 and S69 located at opposite ends of the same β-strand. One possibility is that binding S69 might restrict the dynamics of the strand, which would be then transmitted to the DNA binding domain anchored at the other end.

One other notable difference in the sugar binding contacts provides a means for discriminating inducing from neutral sugars (Table 2, black): Glycerol contact to D149 is either lengthened (WT, 1jyf) or nonexistent (K84L, 1jye), and ligand does not contact L148 or P127 in either structure. Interestingly, TMD indicates that one of the first steps in the overall conformational change (and specifically at the start of pathway 1) is that hydrogen bonds are broken between D149/L128 and D149/Y126 (*42*). Further, the loop containing residue 149 was very dynamic during this simulation, which by necessity did not include coordinates for either sugar ligand (*42*). By not having these contacts, neutral sugar may not be able to restrain the protein in the induced conformation, perhaps reducing an energetic constraint of the induced structure. Although ONPF does contact residue D149 (but interestingly not P127), the induction effects of these contacts could be offset by interactions with the hydrophobic wedge comprising F161, F293, and L296 side chains that prop open the N-subdomain (Figure 5). Similar arguments may apply for L73, P76, and I79, which show close contacts to both IPTG and ONPF (Table 2).

CONCLUSIONS

This paper reports several previously unrecognized differences between functionally distinct LacI states. First, the interface for unliganded LacI is a hybrid of the DNA- and inducer-bound states that are the functional end points of the repressor cycle. Indeed, the interface pattern is very similar to the pattern of cross-monomer interactions that evolve as the protein progresses through the conformational change as simulated by TMD (*42*). Thus the aporepressor appears to be poised to adopt either liganded conformation. Second, the interface of K84L also has a similar intermediate pattern. Combined with results presented in a forthcoming paper that show that this variant is only partially inducible, the conformational change appears to be impeded halfway through the induction process in this LacI variant.

Additional surveys of the sugar binding pocket provide a rationale to discriminate the varied modes of sugar effectors. A number of residues are involved in binding all three sugars and, thus, must not be part of the allosteric communication network. However, anti-inducer ONPF apparently stabilizes a hydrophobic wedge (side chains from F161, F293, and L296) at the back of the binding pocket, "propping" the N-subdomain open in the conformation with high affinity for operator DNA. Further, S69, L148, and D149 appear to be prominent in communicating induction through the repressor structure. Interestingly, the role of D149 was also highlighted by the TMD simulation of the structural shift (*42*), and the L148F substitution has profound effects upon LacI function, presumably through altering the distribution of allosteric conformations (*63*).

Finally, lacking a high-resolution DNA-bound structure in the absence of ONPF, the binding pocket and interface of this biologically relevant structure remain unknown. However, based on the current structures, a reasonable prediction is that the interactions between ~70−79 and the positions of 69−72 C^{α} will look like these regions in unliganded LacI. Interestingly, the low-resolution structure of tetramer bound to DNA but not ONPF (1lbg) (*40*) supports this idea: the C^{α} positions of residues 69−72 are very different than those of the ONPF-bound structures.

Together, these observations provide insight into the subtle structural details that differentiate functionally distinct states

of LacI. Comparative network analysis of these structures was critical for identifying these regions. Similar networks should be generally useful for representing interactions between structurally distinct protein regions, such as inter-subunit, interdomain, or inter-subdomain interfaces, or between protein and macromolecular ligands (e.g., DNA or RNA) (*50*). These networks retain significant structural information, but their visual simplicity facilitates simultaneous comparison of multiple structures, targeting specific regions for detailed structural examination and future experimental tests.

ACKNOWLEDGMENT

I thank Dr. Kathleen Shive Matthews (Rice University) for generous financial support and effective dialogues concerning the manuscript, as well as Dr. Sarah Bondos (Rice University) and Hongli Zhan (Rice University), with whom I had many fruitful discussions. Dr. Jianpeng Ma (Baylor College of Medicine) and Terence Flynn (Rice University) were very helpful in extending the molecular dynamics simulation of the induced crystal structure in the absence of IPTG. Drs. Susan Cates (Rice University) and Jayashree Soman (Rice University) provided assistance in understanding issues related to rmsd calculations and evaluating the quality of X-ray crystallographic structures.

REFERENCES

1. Matthews, K. S., and Nichols, J. C. (1998) Lactose repressor protein: Functional properties and structure, *Prog. Nucleic Acid Res. Mol. Biol. 58*, 127−164.
2. Alberti, S., Oehler, S., von Wilcken-Bergmann, B., Krämer, H., and Müller-Hill, B. (1991) Dimer-to-tetramer assembly of *lac* repressor involves a leucine heptad repeat, *New Biol. 3*, 57−62.
3. Alberti, S., Oehler, S., von Wilcken-Bergmann, B., and Müller-Hill, B. (1993) Genetic analysis of the leucine heptad repeats of Lac repressor: Evidence for a 4-helical bundle, *EMBO J. 12*, 3227−3236.
4. Chakerian, A. E., Tesmer, V. M., Manly, S. P., Brackett, J. K., Lynch, M. J., Hoh, J. T., and Matthews, K. S. (1991) Evidence for leucine zipper motif in lactose repressor protein, *J. Biol. Chem. 266*, 1371−1374.
5. Chen, J., and Matthews, K. S. (1992) Deletion of lactose repressor carboxyl-terminal domain affects tetramer formation, *J. Biol. Chem. 267*, 13843−13850.
6. Chen, J., Surendran, R., Lee, J. C., and Matthews, K. S. (1994) Construction of a dimeric repressor: Dissection of subunit interfaces in *lac* repressor, *Biochemistry 33*, 1234−1241.
7. Chen, J., and Matthews, K. S. (1994) Subunit dissociation affects DNA binding in a dimeric *lac* repressor produced by C-terminal deletion, *Biochemistry 33*, 8728−8735.
8. Bell, C. E., and Lewis, M. (2000) A closer view of the conformation of the *lac* repressor bound to operator, *Nat. Struct. Biol. 7*, 209−214.
9. Bell, C. E., and Lewis, M. (2001) Crystallographic analysis of *lac* repressor bound to natural operator O^1, *J. Mol. Biol. 312*, 921−926.
10. Bell, C. E., Barry, J., Matthews, K. S., and Lewis, M. (2001) Structure of a variant of *lac* repressor with increased thermostability and decreased affinity for operator, *J. Mol. Biol. 313*, 99−109.
11. Barkley, M. D., Riggs, A. D., Jobe, A., and Bourgeois, S. (1975) Interaction of effecting ligands with *lac* repressor and repressor-operator complex, *Biochemistry 14*, 1700−1712.
12. Bartlett, G. J., Borkakoti, N., and Thornton, J. M. (2003) Catalysing new reactions during evolution: Economy of residues and mechanism, *J. Mol. Biol. 331*, 829−860.
13. Luscombe, N. M., and Thornton, J. M. (2002) Protein-DNA interactions: Amino acid conservation and the effects of mutations on binding specificity, *J. Mol. Biol. 320*, 991−1009.
14. Fygenson, D. K., Needleman, D. J., and Sneppen, K. (2004) Variability-based sequence alignment identifies residues responsible for functional differences in α and β tubulin, *Protein Sci. 13*, 25−31.
15. Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol. 3*, 318−356.
16. Gilbert, W., and Müller-Hill, B. (1967) The *lac* operator is DNA, *Proc. Natl. Acad. Sci. U.S.A. 58*, 2415−2421.
17. Gilbert, W., and Müller-Hill, B. (1966) Isolation of the *lac* repressor, *Proc. Natl. Acad. Sci. U.S.A. 56*, 1891−1898.
18. Suckow, J., Markiewicz, P., Kleina, L. G., Miller, J., Kisters-Woike, B., and Müller-Hill, B. (1996) Genetic studies of the *lac* repressor XV†: 4000 Single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure, *J. Mol. Biol. 261*, 509−523.
19. Miller, J. H., Ganem, D., Lu, P., and Schmitz, A. (1977) Genetic studies of the *lac* repressor I. Correlation of mutational sites with specific amino acids residues: construction of a collinear gene-protein map, *J. Mol. Biol. 109*, 275−301.
20. Chakerian, A. E., Olson, J. S., and Matthews, K. S. (1987) Thermodynamic analysis of inducer binding to the lactose repressor protein: Contributions of galactosyl hydroxyl groups and beta substituents, *Biochemistry 26*, 7250−7255.
21. Chakerian, A. E., and Matthews, K. S. (1988) Regulation of the lactose repressor, *Int. J. Biochem. 20*, 493−498.
22. Daly, T. J., and Matthews, K. S. (1986) Allosteric regulation of inducer and operator binding to the lactose repressor, *Biochemistry 25*, 5479−5484.
23. Friedman, B. E., Olson, J. S., and Matthews, K. S. (1976) Kinetic studies of inducer binding to lactose repressor protein, *J. Biol. Chem. 251*, 1171−1174.
24. O'Gorman, R. B., and Matthews, K. S. (1977) Fluorescence and ultraviolet spectral studies of *lac* repressor modified with N-bromosuccinimide, *J. Biol. Chem. 252*, 3572−3577.
25. O'Gorman, R. B., Dunaway, M., and Matthews, K. S. (1980) DNA binding characteristics of lactose repressor and the trypsin-resistant core repressor, *J. Biol. Chem. 255*, 10100−10106.
26. O'Gorman, R. B., Rosenberg, J. M., Kallai, O. B., Dickerson, R. E., Itakura, K., Riggs, A. D., and Matthews, K. S. (1980) Equilibrium binding of inducer to *lac* repressor-operator DNA complex, *J. Biol. Chem. 255*, 10107−10114.
27. Whitson, P. A., and Matthews, K. S. (1986) Dissociation of the lactose repressor-operator DNA complex: Effects of size and sequence context of operator-containing DNA, *Biochemistry 25*, 3845−3852.
28. Whitson, P. A., Olson, J. S., and Matthews, K. S. (1986) Thermodynamic analysis of the lactose repressor-operator DNA interaction, *Biochemistry 25*, 3852−3858.
29. Riggs, A. D., Newby, R. F., and Bourgeois, S. (1970) *lac* repressor-operator interaction II. Effect of galactosides and other ligands, *J. Mol. Biol. 51*, 303−314.
30. Jobe, A., and Bourgeois, S. (1972) *lac* repressor-operator interaction. VI. The natural inducer of the *lac* operon, *J. Mol. Biol. 69*, 397−408.
31. Daly, T. J., and Matthews, K. S. (1986) Characterization and modification of a monomeric mutant of the lactose repressor protein, *Biochemistry 25*, 5474−5478.
32. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., and Takeda, Y. (1982) Structure of the DNA-binding region of *lac* repressor inferred from its homology with *cro* repressor, *Proc. Natl. Acad. Sci. U.S.A. 79*, 1428−1432.
33. Nichols, J. C., Vyas, N. K., Quiocho, F. A., and Matthews, K. S. (1993) Model of lactose repressor core based on alignment with sugar-binding proteins is concordant with genetic and chemical data, *J. Biol. Chem. 268*, 17602−17612.
34. Beyreuther, K., Adler, K., Fanning, E., Murray, C., Klemm, A., and Geisler, N. (1975) Amino acid sequence of *lac* repressor from *Escherichia coli*: Isolation, sequence analysis and sequence assembly of tryptic peptides and cyanogen-bromide fragments, *Eur. J. Biochem. 59*, 491−509.
35. Friedman, A. M., Fischmann, T. O., and Steitz, T. A. (1995) Crystal structure of *lac* repressor core tetramer and its implications for DNA looping, *Science 268*, 1721−1727.
36. Chuprina, V. P., Rullmann, J. A. C., Lamerichs, R. M. J. N., Vanboom, J. H., Boelens, R., and Kaptein, R. (1993) Structure of the complex of *lac* repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics, *J. Mol. Biol. 234*, 446−462.
37. Slijper, M., Bonvin, A. M. J. J., Boelens, R., and Kaptein, R. (1996) Refined structure of *lac* repressor headpiece (1−56) determined by relaxation matrix calculations from 2D and 3D NOE data:

Change of tertiary structure upon binding to the *lac* operator, *J. Mol. Biol. 259*, 761−773.
38. Kalodimos, C. G., Bonvin, A. M., Salinas, R. K., Wechselberger, R., Boelens, R., and Kaptein, R. (2002) Plasticity in protein-DNA recognition: *lac* repressor interacts with its natural operator O^1 through alternative conformations of its DNA-binding domain, *EMBO J. 21*, 2866−2876.
39. Spronk, C. A., Bonvin, A. M., Radha, P. K., Melacini, G., Boelens, R., and Kaptein, R. (1999) The solution structure of *lac* repressor headpiece 62 complexed to a symmetrical *lac* operator, *Struct. Folding Des. 7*, 1483−1492.
40. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer, *Science 271*, 1247−1254.
41. Swint-Kruse, L., Elam, C. R., Lin, J. W., Wycuff, D. R., and Matthews, K. S. (2001) Plasticity of quaternary structure: Twenty-two ways to form a LacI dimer, *Protein Sci. 10*, 262−276.
42. Flynn, T. C., Swint-Kruse, L., Kong, Y., Booth, C., Matthews, K. S., and Ma, J. (2003) Allosteric transition pathways in the lactose repressor protein core domains: Asymmetric motions in a symmetric homodimer, *Protein Sci. 12*, 2523−2541.
43. Matthews, K. S., Matthews, H. R., Theilmann, H. W., and Jardetzky, O. (1973) Ultraviolet difference spectra of the lactose repressor protein, *Biochim. Biophys. Acta 295*, 159−165.
44. Whitson, P. A., Burgum, A. A., and Matthews, K. S. (1984) Trinitrobenzenesulfonate modification of the lysine residues in lactose repressor protein, *Biochemistry 23*, 6046−6052.
45. Alexander, M. E., Burgum, A. A., Noall, R. A., Shaw, M. D., and Matthews, K. S. (1977) Modification of tyrosine residues of the lactose repressor protein, *Biochim. Biophys. Acta 493*, 367−379.
46. Spronk, C. A. E. M., Slijper, M., van Boom, J. H., Kaptein, R., and Boelens, R. (1996) Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator, *Nat. Struct. Biol. 3*, 916−919.
47. Chang, W. I., Olson, J. S., and Matthews, K. S. (1993) Lysine 84 is at the subunit interface of lac repressor protein, *J. Biol. Chem. 268*, 17613−17622.
48. Nichols, J. C., and Matthews, K. S. (1997) Combinatorial mutations of lac repressor: Stability of monomer-monomer interface is increased by apolar substitution at position 84, *J. Biol. Chem. 272*, 18550−18557.
49. Rhodes, G. (2000) *Crystallography made crystal clear*, Academic Press, San Diego, CA.
50. Swint-Kruse, L., Larson, C., Pettitt, B. M., and Matthews, K. S. (2002) Fine-tuning function: Correlation of hinge domain interactions with functional distinctions between LacI and PurR, *Protein Sci. 11*, 778−794.
51. Greene, L. H., and Higman, V. A. (2003) Uncovering network systems within protein structures, *J. Mol. Biol. 334*, 781−791.
52. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) Automated analysis of interatomic contacts in proteins, *Bioinformatics 15*, 327−332.
53. Shindyalov, I. N., and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng. 11*, 739−747.
54. Carson, M. (1997) Ribbons, *Methods Enzymol. 277*, 493−505.
55. Swint-Kruse, L., and Matthews, K. S. (2004) Thermodynamics, protein modification, and molecular dynamics in characterizing lactose repressor protein: Strategies for complex analyses of protein structure−function, *Methods Enzymol. 397*, 188−209.
56. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
57. Carugo, O., and Pongor, S. (2002) Recent progress in protein 3D structure comparison, *Curr. Protein Pept. Sci. 3*, 441−449.
58. Mowbray, S. L., and Bjorkman, A. J. (1999) Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need, *J. Mol. Biol. 294*, 487−499.
59. Chakerian, A. E., Pfahl, M., Olson, J. S., and Matthews, K. S. (1985) A mutant lactose repressor with altered inducer and operator binding parameters, *J. Mol. Biol. 183*, 43−51.
60. Brooks, C. L., III, Karplus, M., and Pettitt, B. M. (1988) Proteins: A theoretical perspective of dynamics, structure, and thermodynamics, in *Advances in Chemical Physics* (Prigogine, I., and Rice, S. A., Eds.) Vol. 71, p 259, John Wiley & Sons, New York.
61. Barry, J. K., and Matthews, K. S. (1997) Ligand-induced conformational changes in lactose repressor: A fluoresence study of single tryptophan mutants, *Biochemistry 36*, 15632−15642.
62. Barry, J. K., and Matthews, K. S. (1999) Substitutions at histidine 74 and aspartate 278 alter ligand binding and allostery in lactose repressor protein, *Biochemistry 38*, 3579−3590.
63. Swint-Kruse, L., Zhan, H., Fairbanks, B. M., Maheshwari, A., and Matthews, K. S. (2003) Perturbation from a distance: Mutations that alter LacI function through long-range effects, *Biochemistry 42*, 14004−14016.
64. Spotts, R. O., Chakerian, A. E., and Matthews, K. S. (1991) Arginine 197 of *lac* repressor contributes significant energy to inducer binding: Confirmation of homology to periplasmic sugar binding proteins, *J. Biol. Chem. 266*, 22998−23002.
65. Gardner, J. A., and Matthews, K. S. (1990) Characterization of two mutant lactose repressor proteins containing single tryptophans, *J. Biol. Chem. 265*, 21061−21067.
66. Chang, W.-I., and Matthews, K. S. (1995) Role of Asp274 in *lac* repressor: Diminished sugar binding and altered conformational effects in mutants, *Biochemistry 34*, 9227−9234.
67. Chang, W.-I., Barrera, P., and Matthews, K. S. (1994) Identification and characterization of aspartate residues that play key roles in the allosteric regulation of a transcription factor: Aspartate 274 is essential for inducer binding in *lac* repressor, *Biochemistry 33*, 3607−3616.

BI049450K