Structure of λ CII: Implications for recognition of direct-repeat DNA by an unusual tetrameric organization

Ajit B. Datta*[†‡], Santosh Panjikar[†], Manfred S. Weiss[†], Pinak Chakrabarti*[§], and Pradeep Parrack*[§]

*Department of Biochemistry, Bose Institute, P1/12 CIT Scheme VIIM, Calcutta 700 054, India; and [†]European Molecular Biology Laboratory, Hamburg Outstation, c/o Deutsches Elektronen Synchrotron, 22603 Hamburg, Germany

Communicated by Allan M. Campbell, Stanford University, Stanford, CA, June 20, 2005 (received for review February 21, 2005)

The temperate coliphage λ, after infecting its host bacterium *Escherichia coli*, can develop either along the lytic or the lysogenic pathway. Crucial to the lysis/lysogeny decision is the homotetrameric transcription-activator protein CII (4 × 11 kDa) of the phage that binds to a unique direct-repeat sequence T-T-G-C-N_6-T-T-G-C at each of the three phage promoters it activates: p_E, p_I, and p_{aQ}. Several regions of CII have been identified for its various functions (DNA binding, oligomerization, and susceptibility to host protease), but the crystal structure of the protein long remained elusive. Here, we present the three-dimensional structure of CII at 2.6-Å resolution. The CII monomer is comprised of four α helices and a disordered C terminus. The first three helices (α1–α3) form a compact domain, whereas the fourth helix (α4) protrudes in different orientations in each subunit. A four-helix bundle, formed by α4 from each subunit, holds the tetramer. The quaternary structure can be described as a dimer of dimers, but the tetramer does not exhibit a closed symmetry. This unusual quaternary arrangement allows the placement of the helix–turn–helix motifs of two of the four CII subunits for interaction with successive major grooves of B-DNA, from one face of DNA. This structure provides a simple explanation for how a homotetrameric protein may recognize a direct-repeat DNA sequence rather than the inverted-repeat sequences of most prokaryotic activators.

direct-repeat recognition | helix–turn–helix motif | transcription activator

The CII protein of phage λ is a transcription activator that has attracted the attention of researchers for diverse reasons. On one hand, CII is a key regulator for the temperate coliphage, enabling the virus to make the choice between the lytic and lysogenic pathways of development (1). On the other hand, CII is the only known example of a DNA-binding protein that recognizes direct-repeat sequences in DNA (2) rather than the commonly observed inverted-repeat sequences in most prokaryotic transcription activators (3).

After infecting its host bacterium *Escherichia coli*, λ can follow either of its two alternate modes of development, namely, lytic or lysogenic. The choice between these pathways is an important decision in the life cycle of the phage (1, 4) and is influenced by factors such as temperature, nutrient concentration, and multiplicity of infection. In fact, this decision-making process was one of the early paradigms for studying gene control and the understanding of molecular switches (5). The phage has developed an intricate molecular machinery for this purpose, involving several proteins from both the phage and the bacterial host (1). Lysogenic development requires the synthesis of λ CI repressor, for which CII is absolutely essential. CII coordinately activates transcription from the three phage promoters p_E (required for initial synthesis of repressor CI), p_I (necessary for synthesizing the integrase protein), and p_{aQ} (directs the synthesis of an antisense RNA to reduce late gene expression), all of which drive the virus toward lysogeny. A striking feature common to all these promoters is the recognition site for CII, a 14-bp direct-repeat sequence T-T-G-C-N_6-T-T-G-C. It has been shown that the four terminal repeat sequences TTGC are essential for CII activity (6, 7). Indeed, CII contacts DNA at the TTGC tetrads (2). Thus, CII represents a unique example of a DNA-binding protein that recognizes direct-repeat, rather than inverted-repeat sequences, as found for most prokaryotic transcription activators (3). Nevertheless, like in other activators (3), it is a helix-turn-helix (HTH) motif that is, presumably, used by CII for DNA recognition (8). The recognition sequence of CII implies that, to contact the TTGC sequences that are positioned 10 bp apart, CII must interact with DNA on one face. How CII achieves this has remained a mystery. The protein could not be crystallized until recently (9), despite its successful purification decades ago (10).

In this article, we report the three-dimensional structure of tetrameric CII at 2.6-Å resolution. It is known that CII exists as a homotetramer in the native state (10), and the tetrameric organization is necessary for its activity and for DNA binding (11). Our crystal structure displays a surprising and unprecedented mode of oligomer formation that makes it possible to explain how the tetrameric molecule may bind to a direct-repeat sequence, as found in the cognate DNA-binding sites of CII. Many biochemical and mutational data on CII are also explained by this structure.

Methods

Protein Expression and Purification. CII was overexpressed and purified from an *E. coli* BL21(DE3) strain transformed with the recombinant plasmid pAB305 according to the protocol described in ref. 9. In short, the protein was overexpressed with the addition of 100 μM isopropyl β-D-thiogalactoside at 37°C to culture at midlog phase for 3 h. Harvested cells were lysed by sonication, and the protein in the soluble phase was purified by 40–55% ammonium sulfate precipitation, followed by two consecutive steps of ion-exchange chromatography using the SP-Sepharose FF (Amersham Pharmacia Biosciences) and High-S (Bio-Rad) columns. The purified protein was dialyzed against 20 mM Tris·HCl (pH 8.0), 1 mM EDTA, and 300 mM NaCl and concentrated to 7 mg/ml. Selenomethionyl derivative was prepared by expressing the protein in *met*$^-$ *E. coli* strain B834(DE3) in the modified M9 media (12), containing selenomethionine at a concentration of 4 mg/ml. The purification protocol was identical to that for the native protein, except for the inclusion of 5 mM β-mercaptoethanol to all of the buffers to prevent the unwanted oxidation of selenomethionyl residues.

Crystallization. Crystals of CII were obtained in hanging drops at 25°C by mixing 2 μl of protein solution and 1 μl of a precipitant

Abbreviations: HTH, helix–turn–helix; MAD, multiple-wavelength anomalous dispersion.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1XWR).

[‡]Present address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14850.

[§]To whom correspondence may be addressed. E-mail: pradeep@bic.boseinst.ernet.in or pinak@bic.boseinst.ernet.in.

Table 1. Data collection, phasing, and refinement statistics

Data collection and phasing	Selenomethionyl				Native
	Peak	Inflection point	Low-energy remote	High-energy remote	
No. of crystals			1		1
Total rotation range, °	110	110	110	180	180
Wavelength, Å	0.9836	0.9841	0.9949	0.9733	0.8013
Space group			$C222_1$		
Cell dimensions (*a, b, c*), Å			64.25, 107.79, 121.01		63.91, 106.79, 119.78
Resolution range, Å			32.57–3.04 (3.09–3.04)		20.0–2.55 (2.59–2.55)
Mosaicity, °	0.46	0.44	0.45	0.44	0.56
Measured reflections	70,959	71,573	71,395	71,307	88,801
Unique reflections	8,391	8,391	8,385	8,388	13,650
Average redundancy	8.5	8.5	8.5	8.5	6.5
$\langle I\rangle/\langle\sigma(I)\rangle$*	9.9 (1.2)	8.7 (1.5)	12.0 (1.4)	12.1 (1.5)	22.4 (2.1)
Completeness, %*	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.8 (99.4)
R_{merge}, %*†	9.5 (68.7)	7.1 (51.8)	6.3 (51.8)	7.6 (50.3)	6.8 (71.9)
R_{rim}, %*‡	10.1 (73.1)	7.6 (55.1)	6.8 (55.1)	8.1 (53.5)	7.5 (78.6)
R_{pim}, %*§	3.4 (24.7)	2.6 (18.5)	2.3 (18.6)	2.7 (18.0)	2.9 (31.3)
R_{anom}, %*¶	6.3 (19.2)	4.0 (17.0)	2.2 (17.5)	4.9 (16.3)	ND
Overall B factor, Å^2	85.5	84.6	87.6	85.8	77.2
Phasing power (acentric)‖	1.58	0	2.18	2.3	—
R_{cullis} (acentric)**	0.68	1.00	0.59	0.57	—
R_{cullis} (anomalous)**	0.70	0.76	0.97	0.74	—
Refinement					
Resolution limits, Å					20.0–2.56
Data cutoff, F/σ (F)					0.0
Total no. of reflections					13,502
Reflections in working set					12,552
Reflections in test set					950
R-factor, %††					24.6
R_{free}, %‡‡					28.6
No. of protein atoms§§					2,440
No. of solvent atoms§§					42
No. of *i*-propanol atoms§§					8
R.m.s. bond lengths, Å					0.009
R.m.s. bond angles, °					1.3
Ramachandran plot (most-favored), %					90.5

ND, not determined.

*Values in parentheses are for the highest resolution shell.

†$R_{merge} = [\Sigma_{hkl}\,\Sigma_i|\,I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\,\Sigma_i\,I_i(hkl)]$, where $I_i(hkl)$ is the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ its average.

‡$R_{rim} = \Sigma_{hkl}[N/(N\text{-}1)]^{1/2}\,\Sigma_i|\,I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\,\Sigma_i\,I_i(hkl)$ (58).

§$R_{pim} = \Sigma_{hkl}[1/(N\text{-}1)]^{1/2}\,\Sigma_i|\,I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\,\Sigma_i\,I_i(hkl)$ (58).

¶$R_{anom} = [\Sigma_{hkl}|\,\langle I(hkl)\rangle - \langle I(-h\ -k\ -l)\rangle|/\Sigma_{hkl}\,0.5|\,\langle I(hkl)\rangle + \langle I(-h\ -k\ -l)\rangle|]$.

‖Phasing power = rms ($|F_H|/E$), where E is the residual lack of closure.

**$R_{cullis} = \Sigma\|F_{PH,O} \pm F_{P,O}| - F_{H,C}|/\Sigma|\,F_{PH,O} - F_{P,O}\|$.

††$R_{factor} = [\Sigma_{hkl}\|F_{obs}| - k\,|F_{calc}\,\|/\Sigma_{hkl}|F_{obs}|]$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

‡‡R_{free}, same for a test set of reflections not used during refinement.

§§per asymmetric unit.

solution, containing 17% (wt/vol) polyethylene glycol (PEG 3350) and 8% (vol/vol) isopropanol. The crystals typically grew in 7–12 days to a maximum size of 0.5 mm and diffracted x-rays typically, to 3.2-Å resolution.

Data Collection. Diffraction data sets were collected at 100 K from crystals flash-frozen in liquid nitrogen after soaking them in cryoprotectant solution containing 17.5% (vol/vol) glycerol. The multiple-wavelength anomalous dispersion (MAD) data sets were collected at the selenium edge at the beamline BW7A of the European Molecular Biology Laboratory (EMBL). The native data set was collected at the fixed-wavelength beamline X13 of EMBL. All data sets were processed by using the programs DENZO and SCALEPACK (13). Table 1 shows the statistics of the data sets used for phasing and structure solution.

Structure Determination and Refinement. F_A values to 4.0 Å, obtained by using the program XPREP (Bruker-AXS, Karlsruhe, Germany) for the selenomethionyl-containing crystal, were used to solve the selenium substructure (12 of 20 seleniums) with the program SHELXD (14). Phases were obtained to 4.0 Å by a four-wavelength MAD experiment. The phases were extended to 3.04 Å by using the program DM (15) in the CCP4 suite of programs (16). The map was of excellent quality. In each monomer, 58 of 97 residues could be built automatically by using the program RESOLVE (17). Once the orientation of the four monomers was known, noncrystallographic symmetry (NCS) operators for each monomer with respect to the first monomer were calculated and then subjected for a second round of density modification in DM for NCS averaging and phase extension to 2.6 Å. The resulting map allowed the building of the loop regions and a number of side chains in each monomer. When there was no further interpretable density in the experiment map, the model was subjected to refinement by using the program CNS (18). The final model contains 320 amino acid residues and exhibits an R factor of 24% (R_{free} = 28%) (for details,

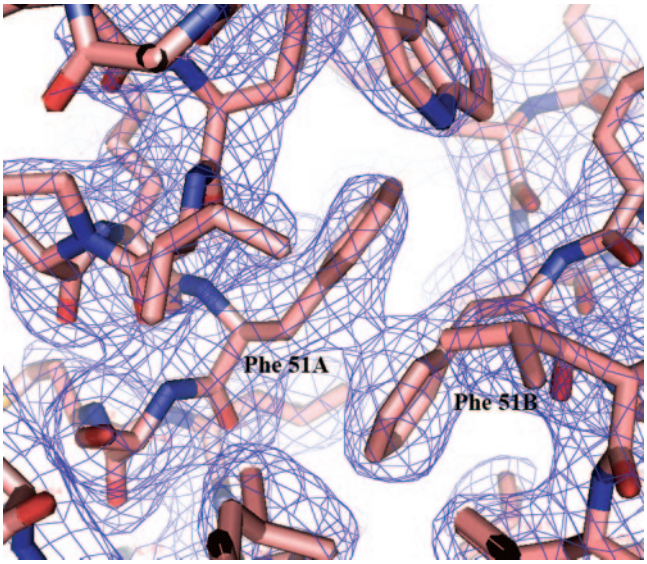

Fig. 1. A section of the $2F_0 - F_C$ map contoured at 1σ, showing the interaction between two Phe rings (belonging to the A and B subunits) across the dimeric interface.

see Table 1). In all four monomers, ≈17 C-terminal residues are disordered. In the Ramachandran plot, 90.5% and 9.5% of all residues are in the most-favored and allowed regions, respectively. The coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 1XWR.

Before collection of the data sets mentioned above, two data sets were also collected at the beamline BL40B2 of SPring-8 (data statistics not shown) (19) to 3.2-Å resolution around the Se edge. Se sites could be located also from the peak data set. However, phasing using either single-wavelength anomalous dispersion (SAD) or the two-wavelength MAD protocol was poor, and, after solvent flattening, the quality of electron density was not sufficient to build the model completely. The poor phasing could be a consequence of radiation damage.

Results and Discussion

Full-length CII crystallizes in the space group $C222_1$, with four molecules in the asymmetric unit (9). Phases were estimated from a MAD experiment on selenomethionine-substituted CII with data extending to 3.04 Å, and the structure was refined against a native data set extending to 2.6 Å (Table 1). A typical section of the electron density map near the dimeric interface AB (see below) is displayed in Fig. 1.

Overall Architecture. The structure of CII tetramer is shown in Fig. 2*A*. It is an all-helix structure in which the four monomers A–D associate via a four-helix bundle near the center of the molecule. The four monomers can be thought of as two pairs (AB and CD). Members of each pair have stronger association between them, compared with that between the pairs. Several aspects of the structure are noteworthy. First, not all of the residues are seen: A significant part of the C terminus is absent in each subunit. Second, the individual monomers are held together in the tetramer predominantly by the C-terminal helix (α4) from each monomer. The orientation of this helix is different in each subunit. Finally, the quaternary structure lacks a closed point group symmetry, an unusual feature for a homotetramer.

A subunit of CII consists of four α-helices, α1–α4, of which the first three form a compact domain (one such subunit is shown in Fig. 2*B*), with the fourth protruding away. Only residues 2–80 were visible in the electron density map: The first (N-terminal) residue and ≈17 residues (this number differs slightly among different subunits) at the C terminus in the polypeptide chain could not be located. This finding indicates a disordered C terminus, as expected from our previous data (20). This flexible C terminus has been found to act as a target for rapid proteolysis by the protease HflB (21) and plays an important role in the lysis/lysogeny decision by making CII unstable *in vivo* (22). Helices α2 and α3 contain residues 26–45, the proposed HTH motif (8) characteristic of prokaryotic regulatory proteins (3, 23). This region of CII has been implicated in DNA-binding from mutational studies (24). The 21-residue helix α3 is longer than the typical second helix in the HTH motif. However, there is actually a break in the helix in subunit C, with the residues Lys-44 to Asp-46 being in a nonhelical conformation. Moreover, the classical HTH motif has a 3-residue long turn, although longer loop regions are also known (25). In CII, this region is 6 residues long. The interhelical angle of ≈90° is typical.

A search against other known folds using the program DALI (26) did not reveal any structural alignment with any other chain of similar length. However, the overall trace of the chain distantly resembles (Z-score = 3.3, rms deviation = 4.7 Å over 52 residues) the 116-residue FlhD (27), which is also a transcriptional activator (of flagellar genes in *E. coli*). In association with FlhC, FlhD acts as a heterotetramer. FlhD itself is dimeric, and the monomer is predominantly α-helical. It is noteworthy that, akin to CII, the N-terminal helix is the longest in the FlhD structure, and the 34 C-terminal residues that protrude from the more compact N-terminal region are mostly disordered.

Structures of CII Monomers in the Tetramer. The most surprising aspect of CII structure is that, although CII is a homotetramer, the orientation of the α-helix 4 (α4) relative to the compact N-terminal domain is different in each of the four subunits (Fig. 2*C*). This conformational plasticity resides in the peptide fragment Glu-59-Trp-60-Gly-61-Val-62-Val-63 between helices α3 and α4. It may be noted that an aromatic residue followed by Gly in this region is conserved among many homologous sequences (11).

Other examples are known of the same amino acid sequence forming two different structures, resulting in an asymmetric subunit interaction. These include the HIV type 1 reverse transcriptase (28), the two-domain transcription regulator cAMP receptor protein from *E. coli* (29), and the LysR-type transcriptional regulator CbnR (30). The latter is a tetrameric structure that can be regarded as a dimer of dimers, where the two subunits in each dimer exist in different conformations, orienting the regulatory domains differently with respect to the linker helix. However, in all three cases, the polypeptide chain has multiple domains with different relative orientations. In contrast, CII is unique because it is a small, single-domain protein, with each of the four chains having a different orientation of the α4 helix relative to the rest of the structure, brought about by conformational flexibility in a short linker region.

Oligomeric Structure and Quaternary Interactions. Homotetrameric proteins generally have the point group symmetry 222 (31). However, the quaternary structure of CII (Fig. 2*A*) does not exhibit any closed point group symmetry. The symmetry axes relating the individual subunits (N-terminal region, residues 7–58) and their pairs are shown in Fig. 2*D*. The individual monomers in the subunit pairs AB and CD are related by twofold axes, implying that the same set of residues in all subunits is involved in dimeric interactions across the interface. A small part of the dimeric interface in AB can be seen in Fig. 1. The two twofold axes in the dimers are not normal to each other as found in most cases, but are inclined at an angle of 28°. Moreover, the dimer CD is related to the dimer AB by a 179° rotation coupled with a translation of 19.3 Å. Such a local screw axis is also known to relate monomers in the hexokinase dimer (32).

The CII tetramer is held together by interactions between the two nearly equivalent dimers AB and CD. The pairwise interface area

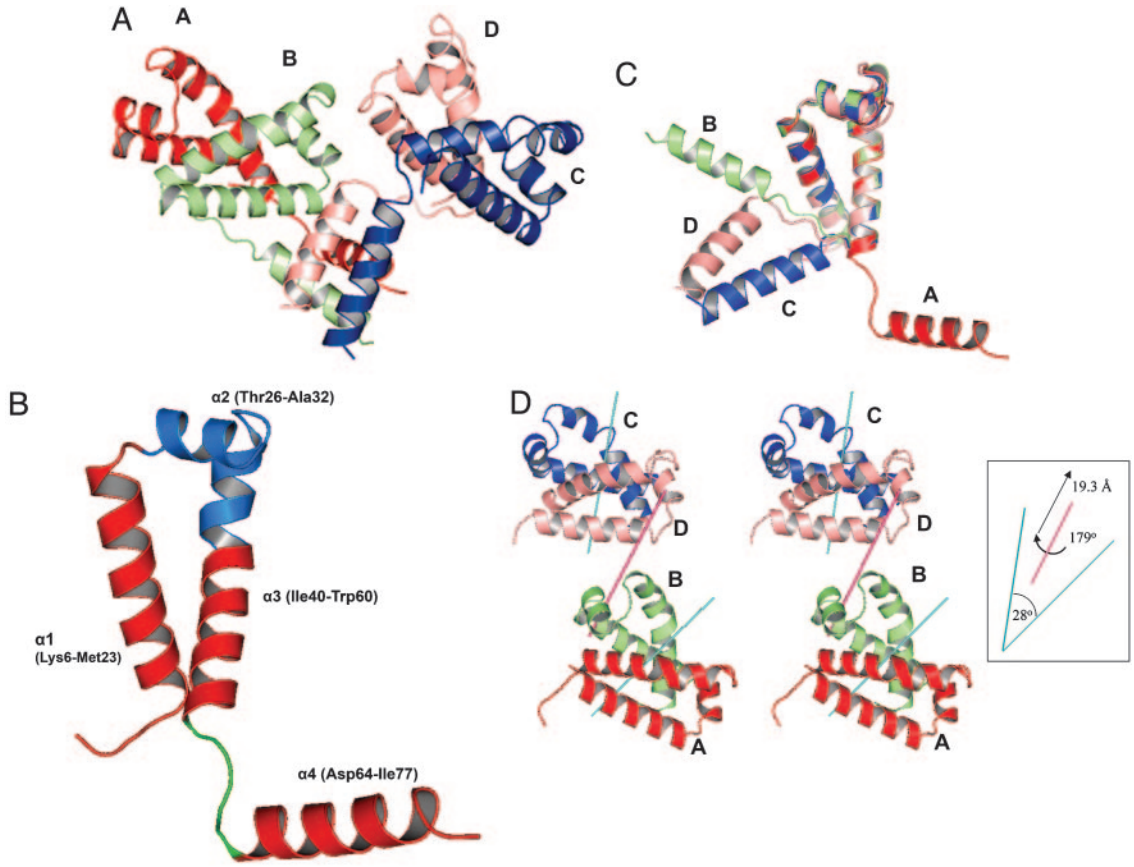

Fig. 2. Structure of the coliphage λ transcription activator protein CII. (*A*) Ribbon representation of the homotetrameric structure. The individual subunits are differently colored and are indicated by letters (A–D). (*B*) Ribbon representation of the structure of one subunit (A chain) of CII. The α-helices and their ranges are marked. The HTH motif is shown in blue, and the flexible loop between α3 and α4 is shown in green. This part has different conformations in the different subunits. (*C*) Superposition of the N-terminal compact domains (α1–α3) of the four subunits (A–D), showing the different orientations of the C-terminal helix α4. (*D*) The local dyads (in cyan within dimers and magenta between dimers) relating the N-terminal domains of the four subunits (A–D). The angular disposition between the two dimeric axes and the rotational and translational components of the screw axis relating the dimers are shown in the box.

buried between the two subunits of these dimers is about the same (2,697 Å^2 and 2,619 Å^2) (Table 2) and is in the same range as that observed for other homodimeric proteins (33, 36). Compared with AB or CD, other subunit pairs (e.g., AC) have a much smaller buried surface. In the AB pair, 68 residues from the two chains constitute the interface, with helices α3 and α4 contributing the most. Among the parameters used to characterize the interfaces (37), the fraction of interface atoms that are completely buried (34%) is comparable to those found in other dimers. However, the number (two) of hydrogen bonds between the chains is lower and the fraction of nonpolar interface area (77%) is higher than the average values (18% and 65%, respectively), indicating that hydrophobic interaction plays a dominant role in holding the CII dimer tightly.

Table 2. Interface areas (Å^2) buried between different pairs of CII subunits

Subunit IDs	B	C	D
A	2,697 (661, 0.24)	400 (400, 1.00)	514 (462, 0.90)
B		196 (196, 1.00)	895 (484, 0.54)
C			2,619 (526, 0.20)

The interface area, ΔASA, represents the surface of the two chains buried when they are brought in contact; it is the sum of the solvent-accessible surface areas (ASA) of the individual subunits minus the ASA of the assembly of the two taken together (33). ASA was calculated by using the program NACCESS (34), which implements the algorithm of Lee and Richards (35). The interface area contributed by the residues in the interacting helices (α4) from the two chains is given in parentheses (the first value denotes area in Å^2, the second denotes the fraction of the total interface area that is contributed by α4).

The dimer–dimer interface formed between AB and CD has an area of 1,692 Å^2, about 60% of the area buried between individual monomers within a dimer. Residues in the α4 helix provide most of the tetrameric contacts, also apparent from the data given in Table 2. For example, the subunits B and D lining the molecular local screw axis bury 895 Å^2 between them, of which 484 Å^2 (i.e., 54%) is accounted for by α4. The importance of α4 (which encompasses residues 64–79) in tetramer formation is expected from our recent data (11), where residues 69–81 were implicated in tetramer formation. Four such helices, one from each subunit of CII, associate to form a four-helix bundle near the center of the molecule (Fig. 2*A*). According to the taxonomy of four-helix-bundle motifs in globular proteins (38), CII corresponds to the type X. This pattern of helix packing is important in the structural organization of several other oligomeric proteins, including two other transcription factors, namely, the tumor-suppressor protein p53 (39) and *lac* repressor (LacI) (40). In the latter, the C-terminal helix (residues 340–357) from each subunit protrudes from the C subdomain and participates in interactions that hold the tetramer, very similar to what we observe in CII. For LacI, deletion of this helix results in a dimeric version that binds a single operator with virtually the same affinity as the tetramer (41). Purine repressor (PurR), another member of the LacI family, lacks the C-terminal

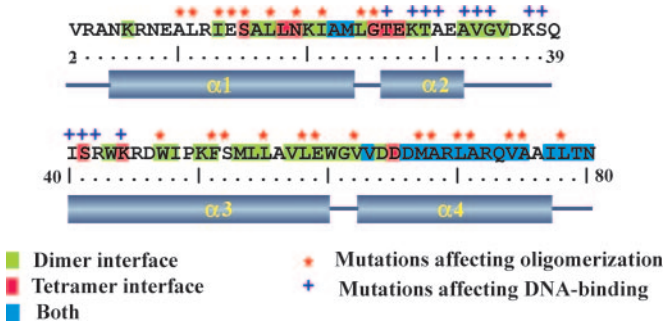

Fig. 3. The sequence of CII (2–80, observed in the structure), with the various structural elements and interface residues highlighted. Mutations that affect oligomerization and DNA-binding of CII are indicated.

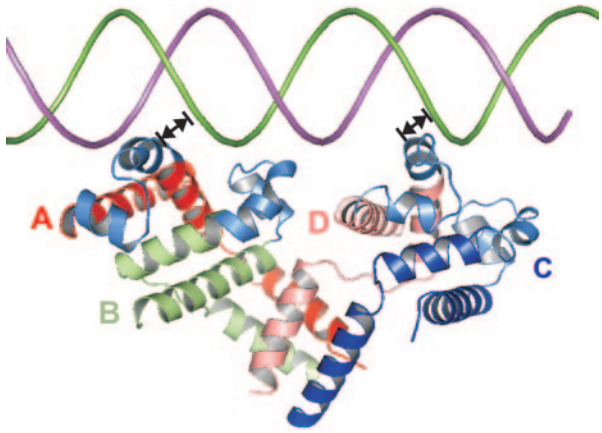

Fig. 4. A model for DNA–CII interaction, showing how the tetrameric assembly can enable recognition of a direct repeat on one face of a B-DNA molecule. All of the HTH motifs are colored blue. The HTH motifs of subunits A and D are located in the major groove of DNA. The HTH from the B subunit may interact with DNA at the minor groove from the same face.

helix and is active as a dimer (42). In comparison, the α4 helices of CII that contribute to most of the tetrameric interactions are indispensable for both DNA binding and the activity of CII (11, 21). Although dimeric mutants of CII have not been found, deletion of the α4 region leads to the formation of dimeric CII *in vitro* (11). These truncated CII molecules, despite retaining the residues important for DNA binding, fail to bind DNA. Thus, tetramer formation is a prerequisite for CII activity.

Another common structural aspect emerges when CII is compared with other homotetrameric DNA-binding proteins. Although the symmetry relating the two dimers in the tetrameric structure is different, CII, lac repressor (40), and CbnR (30) are all roughly V-shaped molecules, with the DNA-binding motifs located on top.

Comparison with Mutational Data. A large number of CII mutants have been studied (2, 24). These data have identified residues in CII that are important for tetramer formation or for DNA binding. Elucidation of the crystal structure enables us to compare these data with our structure. Such a comparison is shown in Fig. 3, where the residues involved in dimer contacts (i.e., between monomers) or between tetramer contacts (between the two dimers) are marked. There is also a set of residues that participate in both dimeric and tetrameric contacts, and these are also indicated separately. A mapping of the reported mutations that affect oligomerization or DNA binding on this schematic diagram shows that (*i*) most mutations that affect oligomer formation are, indeed, located on an interface region (dimer/tetramer/both); the only exceptions are mutations at Ala-10 and Leu-11 that belong to the α1 helix, and (*ii*) mutations that affect DNA binding are located exclusively within the HTH motif formed by α2–α3. This region of the protein is most likely to be involved in DNA recognition, as discussed below. Significantly, mutations that lead to the loss of oligomerization also lead to the loss of activity. Furthermore, no dimeric mutant of CII has been found: Mutations that affect oligomerization invariably give rise to monomeric CII. This finding can be explained from our structure, because most of these mutations are located in the dimeric (or both dimeric and tetrameric) interfaces (here, again, we have exceptions in the form of mutations at Ser-15, Asn-19, and Gly-25 that are involved exclusively in tetramer contacts and not in dimeric contacts).

Direct-Repeat Recognition. Recognition of the direct-repeat sequence T-T-G-C-N_6-T-T-G-C and binding to the same strand of DNA distinguishes CII from other prokaryotic transcription activators. This sequence is conserved in the -35 region of the three λ promoters that CII activates. Presumably, other CII or CII-like proteins from lambdoid phages also have a closely similar sequence for CII binding. It has been shown that the change of one base in either of the tetrad repeats, although resulting in reduced CII binding and impaired lysogenicity, may be tolerated (43). The C1 protein of P22 (which is analogous to λ CII) recognizes the T-T-G-C-N_6-T-T-G-T sequence present in p_E of P22 (44, 45). What makes CII recognize such direct repeats? The x-ray structure presented here offers an explanation for this unusual property of the protein. In the absence of the structure of CII-DNA cocrystal, we constructed a model for CII-DNA interaction by juxtaposition of our CII structure to a B-DNA model (shown in Fig. 4). Placement of the HTH motif of subunit A into the major groove brings the HTH of subunit D very close to the next major groove, with the same DNA strand interacting with the two motifs. This is the only way that two HTH motifs of the protein may contact successive major grooves in B-DNA. The A and D subunits of CII are related by a 38-Å translation and a 28° rotation. This relationship means that the same set of residues from the two subunits are likely to interact with equivalent positions in DNA, 10 bp apart and in the same strand (ideally, this would require a translation of 34 Å and a rotation of 0°). Some distortion of the protein (or DNA) would be required for simultaneous major-groove binding by the two HTH motifs that may be accommodated by conformational changes in the DNA, or in the protein, or in both. The structure of CII, apart from this mismatch of distance (and orientation) of the two HTH motifs, is ideally suited for binding B-DNA from the same face and at successive major grooves, implying recognition of a direct-repeat sequence with a repeat distance of 10 bp, as found in CII-binding sequences.

We note that the recognition helices of the HTH motifs from only the A and D subunits can fit into successive major grooves of DNA. What happens to the HTH belonging to the other two subunits (B and C)? We cannot answer this question, at this point, with much certainty. However, this mode of binding would also bring the HTH motif from subunit B close to the intervening minor groove (in the N6 region) and may allow interactions between its HTH motif and the DNA through the minor groove. Such α-helix–DNA interactions through the minor groove, although known (40, 42), have not hitherto been observed for helices belonging to HTH motifs. However, the existence of such interactions, involving DNA in the region between the repeat sequences, may explain the differences in the affinities of CII for the operator sequences at p_E, p_I, and p_{aQ} (A.B.D. and P.P., unpublished results). The possibility of such additional interactions is also suggested by the behavior of analogous proteins from other lambdoid phages. CII protein from λ or P21 and C1 from P22 is each highly specific for its own p_E promoter (46).

It may also be noted that the model of the CII–DNA complex,

although it explains the phenotypic character of all of the CII mutants, does not explain the striking result obtained with one of its homologous proteins, P22 C1. For this protein, it was shown that a single base change in the N6 region of the promoter (*osi* mutation), which also overlapped with the protein coding sequence and led to Thr-6 being changed to Ala, resulted in an altered specificity. Whereas the wild-type protein could activate transcription from both the wild-type and the mutant promoter (albeit with some difference in the expression level), the mutant protein could activate only the mutant promoter (47), suggesting a possible interaction between the DNA and the N-terminal region of C1. However, for λ CII, we do not find any possible direct interaction between this region from any of the monomers and the promoter DNA. Although the exact reason for this anomaly cannot be ascertained, it seems reasonable to comment that these two proteins, despite their homology, do not interact in the same manner with their respective cognate DNA sites. Such a difference in their modes of interaction has also been suggested (47). The other possibility is that the N-terminal region, although not being important for DNA binding, *per se*, affects a subsequent step in transcription activation, which requires the formation of a ternary complex involving DNA, the activator protein, and the RNA polymerase.

Promoter Activation. An interesting feature of activation by CII is that all its cognate promoters are intrinsically weak, having poor −10 and −35 sequences, and depend on CII binding for their transcription. In fact, mutations that result in a strong −10 sequence can make the p_E promoter a constitutive one, able to function without CII (48). Effective regulation of these promoters perhaps necessitates that they be weak and function only upon the binding of CII. The CII-binding site flanks the −35 region, where the σ4 of RNA polymerase also binds (49). Is it then possible that CII helps the binding of RNA polymerase at the −35 region? It is known that CII and RNA polymerase can simultaneously bind to this part of DNA, at opposite faces (2). The C-terminal domain of the α-subunit of RNA polymerase (αCTD) contacts DNA immediately upstream of the −35 region of p_E, on the opposite face of DNA with respect to CII (50). αCTD is required for the CII-dependent activation of p_E (51, 52) and for lysogenization (53). Interactions between αCTD and DNA (or between αCTD and σ) have been found to be critical for activation of the p_E, p_I, and p_{aQ} promoters by CII, and the possibility of CII–σ interactions has been suggested (54). Synergistic binding of CII and RNA polymerase to DNA has been observed (2, 55); kinetic assays have also shown that CII affects kinetic parameters for RNA polymerase–promoter interactions (56). In the DNA-bound state, CII would be in the immediate vicinity of both the α (49) and the σ subunits (2, 57) of RNA polymerase. All of these considerations strongly suggest contacts between CII and RNA polymerase, although direct evidence for such contacts is not available. Such a ternary CII–DNA–polymerase complex may account for the effective activation at the weak CII-dependent promoters and allow for its control at different levels. Residues from the B and C subunits of CII (Fig. 4) may be involved in such contacts to stabilize the ternary complex.

We thank Prof. Amos Oppenheim for discussions and Dr. Gautam Basu for critical comments on the manuscript. Diffraction data were collected at synchrotron sources at the Deutsches Elektronen Synchrotron and at SPring-8 (Hyogo, Japan). We thank the Department of Science and Technology (India) for supporting travel. This work was supported, in part, by Council of Scientific and Industrial Research (India) Grant 37/1071/01-EMR-II (to P.P.) and a fellowship (to A.B.D.). A.B.D.'s stay in Hamburg was supported by a United Nations Educational, Scientific and Cultural Organization–International Council for Science–Third World Academy of Sciences fellowship.

1. Herskowitz, I. & Hagen, D. (1980) *Annu. Rev. Genet.* **14,** 399–445.
2. Ho, Y.-S., Wulff, D. & Rosenberg, M. (1983) *Nature* **304,** 703–708.
3. Pabo, C. O. & Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61,** 1053–1095.
4. Echols, H. (1986) *Trends Genet.* **2,** 26–30.
5. Ptashne, M. (1992) *A Genetic Switch* (Cell Press, Blackwell Scientific, Cambridge, MA), 2nd Ed.
6. Wulff, D. L., Beher, M., Izumi, S., Beck, J., Mahoney, M., Shimatake, H., Brady, C., Court, D. & Rosenberg, M. (1980) *J. Mol. Biol.* **138,** 209–230.
7. Wulff, D. L., Mahoney, M., Shatzman, A. & Rosenberg, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 555–559.
8. Brennan, R. G. & Matthews, B. W. (1989) *J. Biol. Chem.* **264,** 1903–1906.
9. Datta, A. B., Chakrabarti, P., Subramanya, H. S. & Parrack, P. (2001) *Biochem. Biophys. Res. Commun.* **288,** 997–1000.
10. Ho, Y.-S., Lewis, M. & Rosenberg, M. (1982) *J. Biol. Chem.* **257,** 9128–9134.
11. Datta, A. B., Roy, S. & Parrack, P. (2005) *J. Mol. Biol.* **345,** 315–324.
12. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. (1993) *Nature* **362,** 219–223.
13. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
14. Schneider, T. R. & Sheldrick, G. M. (2002) *Acta Crystallogr. D* **58,** 1772–1779.
15. Cowtan, K. D. & Main, P. (1998) *Acta Crystallogr. D* **54,** 487–493.
16. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D* **50,** 760–763.
17. Terwilliger, T. C. (2002) *Acta Crystallogr. D* **58,** 1937–1940.
18. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
19. Japan Synchrotron Radiation Research Institute (2002) *SPring-8 User Experiment Report No. 9* (Japan Synchrotron Radiation Research Institute, Hyogo, Japan), p. 192.
20. Datta, A. B., Roy, S. & Parrack, P. (2003) *Eur. J. Biochem.* **270,** 4439–4446.
21. Kobiler, O., Koby, S., Teff, D., Court, D. & Oppenheim, A. B. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 14964–14969.
22. Banuett, F., Hoyt, M. A., McFarlane, L., Echols, H. & Herskowitz, I. (1986) *J. Mol. Biol.* **187,** 213–224.
23. Harrison, S. C. (1991) *Nature* **353,** 715–719.
24. Ho, Y.-S., Mahoney, M. E., Wulff, D. L. & Rosenberg, M. (1988) *Genes Dev.* **2,** 184–195.
25. Wintjens, R. & Rooman, M. (1996) *J. Mol. Biol.* **262,** 294–313.
26. Holm, L. & Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138.
27. Campos, A., Zhang, R., Alkire, R. W., Matsumura, P. & Westbrook, E. M. (2001) *Mol. Microbiol.* **39,** 567–580.
28. Wang, J., Smerdon, S. J., Jäger, J., Kohlstaedt, L. A., Rice, P. A., Friedman, J. M. & Steitz, T. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 7242–7246.
29. Weber, I. T. & Steitz, T. A. (1987) *J. Mol. Biol.* **198,** 311–326.
30. Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K. & Senda, T. (2003) *J. Mol. Biol.* **328,** 555–566.
31. Goodsell, D. S. & Olson, A. J. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29,** 105–153.
32. Steitz, T. A., Fletterick, R. J., Anderson, W. F. & Anderson, C. M. (1976) *J. Mol. Biol.* **104,** 197–222.
33. Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. (2003) *Proteins* **53,** 708–719.
34. Hubbard, S. (1992) NACCESS: A Program for Calculating Accessibilities (Univ. Coll. of London, London).
35. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
36. Jones, S. & Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 13–20.
37. Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. (2004) *J. Mol. Biol.* **336,** 943–955.
38. Harris, N. L., Presnell, S. R. & Cohen, F. E. (1994) *J. Mol. Biol.* **236,** 1356–1368.
39. Clore, G. M., Omichinski, J. G., Sakaguchi, K., Zambrano, N., Sakamoto, H., Appella, E. & Gronenborn, A. M. (1994) *Science* **265,** 386–391.
40. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996) *Science* **271,** 1247–1254.
41. Brenowitz, M., Mandal, N., Pickar, A., Jamison, E. & Adhya, S. (1991) *J. Biol. Chem.* **266,** 1281–1288.
42. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. (1994) *Science* **266,** 763–770.
43. Shih, M.-C. & Gussin, G. N. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6432–6436.
44. Backhaus, H. & Petri, J. B. (1984) *Gene* **32,** 289–303.
45. Poteete, A. R., Hehir, K. & Sauer, R. T. (1986) *Biochemistry* **25,** 251–256.
46. Wulff, D. L. & Mahoney, M. E. (1987) *Genetics* **115,** 597–604.
47. Retallack, D. M., Johnson, L. L., Ziegler, S. F., Strauch, M. A. & Friedman, D. I. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9562–9565.
48. Keilty, S. & Rosenberg, M. (1987) *J. Biol. Chem.* **262,** 6389–6395.
49. Dove, S. L., Darst, S. A. & Hochschild, A. (2003) *Mol. Microbiol.* **48,** 863–874.
50. Kedzierska, B., Lee, D. J., Wegrzyn, G., Busby, S. J. W. & Thomas, M. S. (2004) *Nucleic Acids Res.* **32,** 834–841.
51. Gussin, G. N., Olson, C., Igarashi, K. & Ishihama, A. (1992) *J. Bacteriol.* **174,** 5156–5160.
52. Obuchowski, M., Giladi, H., Koby, S., Szalewska-Palsz, A., Wegrzyn, A., Oppenheim, A. B., Thomas, M. S. & Wegrzyn, G. (1997) *Mol. Gen. Genet.* **254,** 304–311.
53. Wegrzyn, G., Glass, R. E. & Thomas, M. S. (1992) *Gene* **122,** 1–7.
54. Marr, M. T., Roberts, J. W., Brown, S. E., Klee, M. & Gussin, G. N. (2004) *Nucleic Acids Res.* **32,** 1083–1090.
55. Shimatake, H. & Rosenberg, M. (1981) *Nature* **292,** 128–132.
56. Shih, M.-C. & Gussin, G. N. (1984) *J. Mol. Biol.* **172,** 489–506.
57. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. (2002) *Mol. Cell* **9,** 527–539.
58. Weiss, M. S. (2001) *J. Appl. Crystallogr.* **34,** 130–135.