BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Binding of the N-Terminal Domain of the Lactococcal Bacteriophage TP901-1 CI Repressor to Its Target DNA: A Crystallography, Small Angle Scattering, and Nuclear Magnetic Resonance Study

Kristian H. Frandsen,[†] Kim K. Rasmussen,[†] Malene Ringkjøbing Jensen,[‡] Karin Hammer,[§] Margit Pedersen,[∥,⊥,@] Jens-Christian N. Poulsen,[†] Lise Arleth,[⊥] and Leila Lo Leggio*[,†]

[†]Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

[‡]Institut de Biologie Structurale Jean-Pierre Ebel, CNRS-CEA-UJF UMR 5075, 38027 Grenoble, France

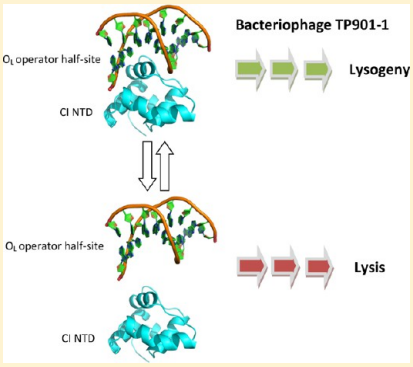
[§]Center for Systems Microbiology, Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark

[∥]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark

[⊥]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Supporting Information

ABSTRACT: In most temperate bacteriophages, regulation of the choice of lysogenic or lytic life cycle is controlled by a CI repressor protein. Inhibition of transcription is dependent on a helix−turn−helix motif, often located in the N-terminal domain (NTD), which binds to specific DNA sequences (operator sites). Here the crystal structure of the NTD of the CI repressor from phage TP901-1 has been determined at 1.6 Å resolution, and at 2.6 Å resolution in complex with a 9 bp double-stranded DNA fragment that constitutes a half-site of the O_L operator. This N-terminal construct, comprising residues 2−74 of the CI repressor, is monomeric in solution as shown by nuclear magnetic resonance (NMR), small angle X-ray scattering, and gel filtration and is monomeric in the crystal structures. The binding interface between the NTD and the half-site in the crystal is very similar to the interface that can be mapped by NMR in solution with a full palindromic site. The interactions seen in the complexes (in the crystal and in solution) explain the observed affinity for the O_R site that is lower than that for the O_L site and the specificity for the recognized DNA sequence in comparison to that for other repressors. Compared with many well-studied phage repressor systems, the NTD from TP901-1 CI has a longer extended scaffolding helix that, interestingly, is strongly conserved in putative repressors of Gram-positive pathogens. On the basis of sequence comparisons, we suggest that these bacteria also possess repressor/antirepressor systems similar to that found in phage TP901-1.



Lactococcus lactis bacteria are used as starter cultures in the dairy industry, and in spite of considerable effort, bacteriophage infections are still a significant problem. The milk in the dairy is not sterile and may contain lytic phages. If the bacterial starter culture is sensitive to these phages, it will be wiped out, and hence, the fermentation will be spoiled. Three groups of phages cause the vast majority of problems, and one of these, the P335 species, also contains temperate bacteriophages, including TP901-1, which is the lactococcal phage characterized best at the molecular level.[1] Transcriptional regulation in temperate bacteriophages, especially the *Escherichia coli* λ phage, has served for decades as a model system for physiological, biochemical, and structural studies, giving important insight into molecular mechanisms that now have become part of the modern paradigm of gene regulation.[2,3]

In general, two alternative life cycles are possible for temperate bacteriophages when infecting a sensitive bacterial host cell. In the lysogenic life cycle, gene expression from the viral genome is repressed, and usually, the genome is integrated into the bacterial chromosome and then duplicated as the host cell proliferates. In the lytic life cycle, the viral genome is transcribed, replicated, and translated followed by assembly to form complete bacteriophages until the host cell is lysed. The choice between these two life cycles, in λ and many bacteriophages, is controlled by a bistable genetic switch. This genetic switch often consists of two divergently oriented promoters from which genes essential for either of the two life cycles are transcribed. In the majority of known temperate bacteriophages, the key player involved in such a switch is the CI (Clear 1) repressor protein. The repressor recognizes various operator sites within the genetic switch and represses transcription from either of two promoters. CI repressor proteins normally consist of an N-terminal domain (NTD) containing a helix−turn−helix (HTH) motif involved in DNA binding followed by a flexible linker region and ending with a

Received: April 8, 2013
Revised: August 27, 2013
Published: September 2, 2013

ACS Publications © 2013 American Chemical Society

C-terminal domain (CTD) responsible for the quaternary structure of the whole repressor. Sequences diverge especially in the CTDs, which are usually unrelated at the sequence level between different phages.

Binding of the NTD to operator sites within the switch with different affinities and the oligomerization of the CTD allow the phages to respond to an adequately potent external stimulus in a cooperative manner and toggle the switch to the lytic cycle. Despite the similarities in the domain structure of CI repressors and general mechanisms, significant differences exist between different phages, in particular with regard to the quaternary organization of CI, and the molecular mechanisms governing the switch from one state to the other.

In the most studied bistable genetic switch (λ phage), the CI repressor is responsible for maintenance of the lysogenic state and can bind to each operator site (O_L1, O_L2, and O_L3 and O_R1, O_R2, and O_R3) as a dimer. Cooperative interactions between CI repressors lead to the formation of tetrameric and octameric states, fine-tuning maintenance of lysogeny and negative autoregulation,[4] so that transcription from lytic promoter P_R is repressed and transcription from lysogenic promoter P_{RM} partially allowed. The two promoter regions are divergently oriented and are arranged back to back. Switching to the lytic state requires the Cro repressor to compete for binding at the DNA operator sites, thus changing the switch to repress transcription from P_{RM} and allow it from P_R.[3] In phage λ, the SOS response triggers self-cleavage and consequent inactivation of CI, removing it from the operators. Cro is then expressed and subsequently binds DNA, inhibiting the synthesis of CI from P_{RM}.

Another well-studied system, the switch of Coliphage 186,[5] shows clear differences at the molecular level from the λ switch. The fully assembled form of 186 CI is a heptamer of dimers (tetradecamer),[5] which binds two types of operator sites termed A and B. These overlap with the lytic promoter region P_R within the switch, while lysogenic promoter P_L is only overlapped by a weak A-type operator site.[5] In contrast to the λ phage switch, these promoters are arranged face to face rather than back to back. The switch of phage 186 furthermore includes two flanking operator sites, F_L and F_R, which also play a role in repression.[5] The switch to the lytic cycle in this phage is also connected to the SOS response through the bacterial regulator protein, LexA, and the phage protein, Tum, which is transcribed from a LexA-controlled operon. Tum functions as an antirepressor and inhibits CI from binding to DNA.[5,6]

In the genetic switch controlling lysogeny in TP901-1, the system under study in this report, the lysogenic state is retained by a hexameric form of CI[7] binding three operator sites (O_L, O_R, and O_D) in a cooperative manner. Apparent dissociation constants of 28 and >2000 nM have been reported for O_L and O_R, respectively, while the affinity for O_D is reported to be similar to that for O_L.[8,9] Full-length CI comprises 180 amino acid residues. Studies of different truncation constructs of CI that removed up to 58 amino acid residues from the C-terminal end have shown that the TP901-1 CTD mediates formation of the full-length CI hexamer, while the construct with 58 C-terminal residues removed was shown to be a dimer.[7] This, supported by SAXS data, has led to a model of full-length CI consisting of a trimer of dimers.[7]

Binding of full-length CI to the three operator sites represses transcription from the lytic promoter, P_L, and allows some transcription from the lysogenic promoter, P_R. Previous mutational studies of CI have indicated the importance of the NTD in DNA binding,[7] but the oligomeric state of this domain in isolation was not investigated.

The lytic state of TP901-1 is also induced by DNA damage, as is the case for phage λ and phage 186. The bacterial RecA protein is also needed for the induction of the TP901-1 prophage, but the mechanism is not known.[10] However, toggling the switch to the lytic cycle involves the phage protein MOR, the modulator of repression, which acts as an antirepressor supposedly by forming a heterodimer with CI and binding a hypothetical operator site, O_M, resulting in repression of the P_R promoter and allowing transcription from P_L.[7,9,11]

This report focuses on the role of the CI-NTD in the binding of DNA and the bistable switch of TP901-1. Constructs bearing the wild-type sequence and ones with the K40A/S41A and Q45A/N48A mutations, previously shown to impair DNA binding,[7] were used for structural analysis to maximize the chances of successful crystallization. We determined the crystal structure of CI-NTDQ45A/N48A at 1.6 Å resolution and the CI-NTD at 2.6 Å resolution in complex with a 9 bp double-stranded DNA (dsDNA) fragment that constitutes a half-site of the O_L operator. The crystal structures are complemented with solution NMR studies of the interaction between the NTD and a full palindromic site. Recent genome projects[12−15] have revealed a number of close homologues of the CI-NTD (>50% identical sequence) in pathogenic Gram-positive bacteria and their phages, broadening the significance of our findings.

MATERIALS AND METHODS

Plasmids and Expression. Constructs were based on expression vector pQE-70 containing sequences encoding the wild-type CI-NTD, or variants containing two amino acid substitutions, CI-NTDK40A/S41A and CI-NTDQ45A/N48A. Expressed proteins consisted of the N-terminal domain (residues 2−74) of the TP901-1 CI repressor protein (or variant thereof), with two additional N-terminal residues originating from the cloning site and a C-terminal six-His tag (-RSHHHHH), for a total of 83 residues. Protein was expressed in *E. coli* strain M15[pREP4] transformed with the pQE-70 vector under the control of a T7 promoter. Correctly transformed cells have both ampicillin and kanamycin resistance. An overnight culture at 37 °C (100 μg/mL ampicillin and 25 μg/mL kanamycin) in 10 mL of LB medium was used to inoculate 1 L of LB medium (or in minimal medium M9 for labeled protein). The culture was grown to an OD_{600} of 0.6 AU, at which point protein expression was induced by adding IPTG to a final concentration of 1 mM. After incubation at 25 °C for approximately 20 h, cells were harvested by centrifugation and the cell pellet was stored at −20 °C for 15 min or until it was used. Resuspended cells were sonicated on ice in 40 mL of buffer A [20 mM Tris-HCl (pH 8), 20 mM imidazole, and 1 M NaCl], centrifuged to remove cell debris, and filtered through a 0.22 μm filter.

Purification. The cell lysate was first applied to a Ni-NTA column in buffer A and washed with buffer A, and the His-tagged protein was then eluted with buffer B [20 mM Tris-HCl (pH 8), 250 mM imidazole, and 1 M NaCl] in a volume of approximately 3 mL. The eluate was loaded onto a HiLoad 26/60 Superdex75 column on an ÄKTA purifier, from which fractions of 5 mL were collected. Fractions were concentrated as needed for later studies using Amicon Ultra Centrifugal filters with a molecular mass cutoff of 3 kDa.

dx.doi.org/10.1021/bi400439y | *Biochemistry* 2013, 52, 6892−6904

Basic Characterization. The protein concentration was estimated by A_{280} using an extinction coefficient of 8480 M^{-1} cm^{-1} as estimated with ProtParam (http://web.expasy.org/protparam/). Protein purity was estimated on 15% sodium dodecyl sulfate−polyacrylamide gel electrophoresis gels stained with InstantBlue (Expedeon). Protein mass was determined by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Bruker AutoFLEX apparatus (Bruker Daltronik GmbH, Bremen, Germany) in positive polarity mode, calibrated with a protein mass MALDI-MS calibration kit (Sigma-Aldrich, MSCAL3-KT). Protein samples were applied by use of the "sandwich method" with sinapinic acid as the matrix as described by Sigma-Aldrich. Mass spectrometry data were analyzed with FlexAnalysis version 2.0.

Oligonucleotides. Oligonucleotides were purchased from TAG Copenhagen (1 μmol scale, high-performance liquid chromatography-purified) and annealed by being incubated for 5 min at 95 °C and then cooled to room temperature. For NMR titrations, an 18 bp dsDNA fragment corresponding to the O_L site[8] was used (5′-GTTCATGAAACGTGAACT-3′ and its complementary strand). A 9 bp dsDNA fragment (5′-AGTTCACGT-3′ and its complementary strand) corresponding to an O_L half-site was used for crystallization.

Crystallization and Cocrystallization. Vapor diffusion crystallization screening was conducted with an Oryx 8 Protein Crystallization Robot and MRC 2 Well Crystallization plates (Douglas Instruments Ltd.) at room temperature using a JCSG + screen (Qiagen) for NTD crystallization or a Natrix screen (Hampton research) for cocrystallization of the CI-NTD with the 9 bp dsDNA fragment. Each well had a reservoir of a 100 μL and two sitting drops of 0.3 μL with reservoir concentrations of 25 and 50%, respectively. Drops were set up using a stock protein concentration of 12 mg/mL. Best crystals for the apo CI-NTD domain were obtained for CI-NTDQ45A/N48A, for which an initial crystal hit was obtained with 1.6 M trisodium citrate as the precipitant. Optimization of the condition in VDX plates (to 1.1 M trisodium citrate as the precipitant), with protein concentrations of 6−12 mg/mL and a drop size of 2 μL, led to diffracting crystals for which the final data set was collected. Because the CI-NTDQ45A/N48A variant is impaired in DNA binding, only the wild-type CI-NTD was used for cocrystallization experiments. The CI-NTD was mixed in a 1:1 molar ratio with the 9 bp dsDNA fragment. Within 4 days, fine needle-shaped crystals were obtained from a screen condition with a reservoir of 50 mM HEPES (pH 7.0), 25% (v/v) PEG 550, and 5 mM $MgCl_2$.

Crystallographic Data Collection and Processing. Crystals were mounted in nylon cryo-loops from Hampton research without the use of a cryoprotectant and data collected at 100 K. Several crystals were tested at beamlines 911-2 and 911-3 of MAXLAB (Lund, Sweden). Finally, diffraction data for an apo crystal of CI-NTDQ45A/N48A were collected at beamline ID29 of the European Synchrotron Research Facility (ESRF, Grenoble, France). Scaling and processing using the XDS package[16] determined that the crystal belonged to orthorhombic space group $P2_12_12_1$ with one molecule in the asymmetric unit and good scaling statistics up to 1.6 Å resolution (Table 1). Data for the complex with DNA were collected at beamline ID23-2 of the ESRF and data processed up to 2.6 Å resolution using MOSFLM.[17] The space group was also determined to be $P2_12_12_1$ but with an unrelated cell. Data collection details and statistics are listed in Table 1.

Table 1. Crystallographic Parameters and Data[a]

	CI-NTDQ45A/N48A	CI-NTD, half-site O_{L2}
beamline (ESRF)	ID29	ID23-2
exposure (s)	0.10	1
oscillation (deg)	0.10	1.0
no. of images	1450	120
detector	Pilatus 6m detector	Mar 225
transmission (%)	2.9	100
unit cell		
a, *b*, *c* (Å)	29.66, 43.81, 72.55	29.86, 64.07, 67.82
α, *β*, *γ* (deg)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
λ (Å)	0.97625	0.87
space group	$P2_12_12_1$	$P2_12_12_1$
resolution limit (Å)	18.60−1.6 (1.7−1.6)	33.91−2.60 (2.74−2.60)
total no. of reflections	65211 (10813)	19294 (2930)
no. of unique reflections	12874 (2105)	4365 (640)
completeness (%)	98.7 (99.5)	100 (100)
I/σ	26.49 (4.90)	9.0 (4.7)
R_{meas} (%)	3.6 (34.4)	16.2 (48.8)
R_{work} (%)	13.51	19.9
R_{free} (%)	18.26	21.9
rmsd for bonds (Å)	0.010	0.002
rmsd for angles (deg)	1.164	0.670
Wilson *B* ($Å^2$)	20.35	22.4
Ramachandran plot (%)		
favored	71/71 (100%)	78/78 (100%)
poor rotamers	1/65 (1.5%)	2/73 (2.74%)
MolProbity score	1.27 (97th percentile)	1.52 (100th percentile)
PDB entry	3zhi	3zhm

[a]Numbers in parentheses refer to data for the outer resolution shell.

Crystal Structure Determination, Refinement, and Validation. A threading search using the PHYRE server[18] listed several structures with a low level of sequence identity (~20%) that were used as search models for determining the structure of CI-NTDQ45A/N48A by molecular replacement (MR) in PHENIX.PHASER.[19,20] Initially, these attempts were unsuccessful, also with side chains pruned and truncated N-terminal domains, probably because of the low level of sequence identity of the models. PHENIX.SCULPTOR[21] was then used to modify the structures of several of these weakly homologous domains according to the CI-NTD sequence. The modified structure of the NTD from the bacteriophage P22 C2 repressor[22] (PDB entry 1ADR, 18% sequence identity with the CI-NTD) was the most successful search model in PHASER. The solution from PHASER was used in PHENIX.AUTOBUILD,[23] which automatically partially built the structure of CI-NTDQ45A/N48A (66 of 83 residues), even though the rotation and translation *Z* scores as well as Log-Likelihood Gain (LLG) were not significantly better than with other search models (RFZ = 3.9, TFZ = 6.1, and LLG = 29). Missing residues could be easily built manually using COOT[24,25] and the structure refined anisotropically in PHENIX.REFINE.[26]

MR was conducted on the DNA complex data by searching with two components, the CI-NTDQ45A/N48A structure determined above and the half-operator site DNA (base pairs 1−9) from the P22 C2 operator complex (PDB entry 2R1J).[22] Initially, only the protein component was found, but evidence of additional density was present in the maps. With the protein

dx.doi.org/10.1021/bi400439y | *Biochemistry* 2013, 52, 6892−6904

Table 2. SAXS Parameters for the CI-NTD (wild type) and BSA (reference)[a]

sample	$R_{g,Guiner}$ (nm)	$R_{g,Gnom}$ (nm)	V_p (nm^3)	D_{max} (nm)	$M_{wExpPorod}$ (kDa)	$M_{wExpI(0)}$ (kDa)	M_{wTheo} (kDa)
BSA	3.0	3.1	114.5	11.1	71.6	66.4	66.4
NTD	1.3	1.3	9.9	4.7	6.2	7.4	9.6

[a]Radii of gyration (R_g) have been estimated either from $I(0)$ determined by the Guinier approximation or from the full scattering curve by GNOM processing. The molecular mass was estimated either from $I(0)$ using BSA as a standard or from the Porod volume (V_p).

placed and used as a partial solution, another round of MR using a truncated model (base pairs 2−9 of the dsDNA from PDB entry 2R1J) as a search model yielded the following scores: RFZ = 3.9, TFZ = 9.7, and LLG = 156. A difference map revealed the omitted base pairs expected in the structure. Refinement was conducted in PHENIX.REFINE[26] alternated with manual rebuilding in COOT.[24,25] Refinement and validation statistics for both structures are listed in Table 1.

Solution X-ray Scattering and Modeling. Small angle X-ray scattering (SAXS) data were collected at station ID14-3 of the ESRF at a wavelength of 0.931 Å. Protein concentrations of 5.1, 5.3, and 7.0 mg/mL were used for CI-NTDK40A/S41A, CI-NTDQ45A/N48A, and CI-NTD, respectively, while a protein concentration of 4.02 mg/mL was used for bovine serum albumin, which was used a reference protein for molecular mass estimation.[27] Ten frames were collected with a constant flow of sample to reduce radiation damage. Data processing was performed using PRIMUS,[28] and GNOM[29] was used to determine the radius of gyration, the pair-distance distribution function $P(r)$, and the maximal particle distance (D_{max}). CRYSOL[30] was used to compare scattering profiles calculated from high-resolution models with the experimental SAXS data. Residues missing in the crystal structure were modeled in extended conformation in COOT.[24,25] Various SAXS-derived parameters are listed in Table 2.

NMR Spectral Assignment of the CI-NTD. Spectral assignment of the CI-NTD was conducted at 25 °C using a 0.7 mM sample at pH 6.5 (20 mM Tris buffer and 100 mM NaCl). The assignment of the CI-NTD was obtained at a ^{1}H frequency of 600 MHz using BEST-type triple-resonance experiments HNCO, intraresidue HN(CA)CO, intraresidue HNCA, HN(CO)CA, and HN(COCA)CB.[31] The spectra were acquired with a sweep width of 8 kHz and 512 complex points in the ^{1}H dimension and a sweep width of 1.4 kHz and 34 complex points in the ^{15}N dimension. For the ^{13}C dimension, the spectra were acquired with a sweep width of 1.4 kHz and 60 complex points [HNCO and intraresidue HN(CA)CO], a sweep width of 3.7 kHz and 80 complex points [HN(CO)CA and intraresidue HNCA], and a sweep width of 10 kHz and 100 complex points [HN(COCA)CB]. The spectra were processed using NMRPipe[32] and analyzed using SPARKY.[33] Automatic assignment of spin systems was conducted using MARS.[34] Two standard ^{1}H−^{15}N HSQC experiments were conducted on the apo CI-NTD at either 350 or 800 μM.

Interaction between the CI-NTD and DNA in Solution. Chemical shift titration experiments were conducted using the ^{15}N- and ^{13}C-labeled CI-NTD with the 18 bp dsDNA fragment. Increasing amounts of DNA were added to the protein, and for each step of the titration, a ^{1}H−^{15}N TROSY spectrum[35] was recorded at a ^{1}H frequency of 800 MHz and 25 °C. Initially, a titration was attempted at a sodium chloride concentration of 100 mM, leading to precipitation in the NMR tube of the protein−DNA complex at the applied concentrations. A new titration was therefore conducted at 300 mM salt where we saw no signs of precipitation, leading to shifting CI-NTD peaks according to a fast to intermediate exchange regime. TROSY spectra were recorded for the following concentrations of CI-NTD and DNA: 284 and 0.0 μM, 271 and 20.6 μM, 259 and 39.3 μM, 246 and 59.8 μM, 231 and 84.1 μM, 211 and 115 μM, 189 and 153 μM, and 166 and 192 μM, respectively (reaching a 1:1.2 concentration ratio at the final titration point).

RESULTS

Protein Production, Mass Spectrometry, and Gel Filtration. The CI-NTD yield (wild-type or mutant) was approximately 10 mg/L of culture. The presence of the correct proteins was verified by MALDI-TOF, which showed masses similar to those calculated from sequence. For example, for CI-NTDQ45A/N48A, the experimental mass was 9513 ± 5 Da, compared to a theoretical value of 9508 Da. When run on gel filtration (Superdex75), the protein (wild type or variant) eluted at a volume (V_E) corresponding to apparent molecular masses of approximately 6.3−6.5 kDa (depending on the variant) estimated by the use of the following standards: ovalbumin (43 kDa), ribonuclease A (13.7 kDa), and aprotinin (6.5 kDa). We conclude that under gel filtration conditions, all investigated forms of the CI-NTD are monomers.

Crystal Structure of CI-NTDQ45A/N48A. Remarkably, given the low level of sequence identity of the search models available, the crystal structure of CI-NTDQ45A/N48A could be determined by MR through careful model preparation. Although MR structure determination problems cannot be compared directly, it is interesting to see that in the case of the structurally related EspR repressor,[36] MR did not succeed, showing that MR at this level of search model similarity is not trivial. It must though be noted that in the case of EspR, twinning and multiple copies in the asymmetric unit provided additional challenges. According to its structure and sequence, the CI-NTD belongs to the Cro and repressor family of HTH transcription factors[37] or the cI-like family in the tetrahelical HTH superclass, as more recently defined in ref 38. The structure of CI-NTDQ45A/N48A (Figure 1) contains five α-helices; the first four are the same as defined for the tetrahelical HTH superclass. The first α-helix of the structure followed by a

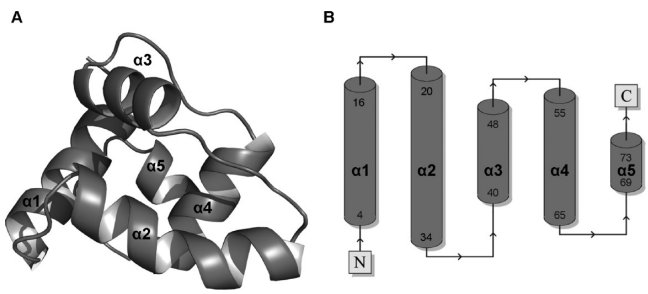


Figure 1. (A) Structure of CI-NTDQ45A/N48A. The five-α-helix fold is shown as a cartoon. The two helices, α2 and α3, constitute the HTH motif. (B) Secondary structure topology diagram with residue numbers and α-helix numbers in bold.

dx.doi.org/10.1021/bi400439y | Biochemistry 2013, 52, 6892−6904

short loop (α1−loop 1) comprises residues Asp4−Lys19. The following residues, Gln20−Asn48, comprise the second and third helices separated by a short loop (α2−loop 2−α3), which constitute the HTH motif, α2 being the scaffolding and α3 the recognition helix (RH). The last part of the structure (Ser49−Gly74) comprises a small peptide segment (loop 3) followed by two helices separated by a short loop (α4−loop 4−α5).

Overall, the structure resembles the mainly α-helical fold of other members of the Cro and repressor family, comprising repressor proteins from other bacteriophages such as CI from λ, 434, and 186 phages,[2,4,5,39] C2 from P22 phage,[22] etc. However, the HTH motif in the CI-NTD is noncanonical, as described more in detail in the Discussion, because the connection between α2 and α3 is not a four-residue turn as in the traditional definition of the HTH motif.[37] CI-NTDQ45A/N48A is monomeric in the crystal structure, in the sense that no extensive contacts exist between symmetry-related molecules.

Small Angle X-ray Scattering. Scattering curves recorded on the CI-NTD (wild type or mutants) were similar, and further analysis was conducted on the wild-type CI-NTD. From the extrapolation to $I(0)$ and using BSA as a standard, the molecular mass of the CI-NTD was found to be 6−7 kDa compared to the theoretical M_w of 9.6 kDa, indicating that the CI-NTD in solution is monomeric at the sole concentration tested for each variant. Significant deviations were observed between the experimental scattering curve of the CI-NTD and that back-calculated on the basis of the crystal structure using CRYSOL (comparison not shown). However, when the missing C-terminal (RSHHHHHH) and N-terminal (ML) residues are modeled in extended conformations onto the crystal structure of the NTD, the model fits well to the experimental data (Figure 2).

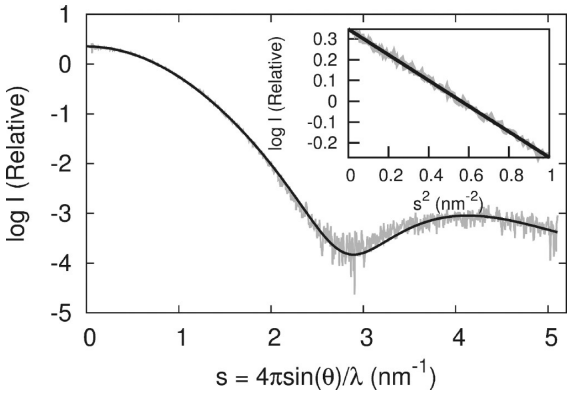


Figure 2. Comparison of experimental (gray) and calculated (black) SAXS scattering curves for the CI-NTD. CRYSOL[30] was used for calculation of the scattering profile from coordinates. The model used to calculate the scattering profile was the crystal structure of CI-NTDQ45A/N48A extended with the missing residues (see Materials and Methods). The inset shows the fit in the Guinier region (linear region expected for nonaggregated samples).

Crystal Structure of the CI-NTD−DNA Complex. The crystal structure of the CI-NTD in complex with a half-operator site reveals that the recognition helix is inserted into the major groove, and the scaffolding helix is positioned across the major groove to bridge the phosphates at either side (Figure 3A,B). The crystal structure of the CI-NTD in the protein−DNA complex and the crystal structure of CI-NTDQ45A/N48A alone are essentially identical except for a few side chain reorganizations. In the DNA complex, Lys40 and Gln51 move slightly as they bind DNA (Gln51 also then interacts with Gln45), while Glu15 interacts with the His tag. The other residues with side chain differences are polar residues Lys2, Lys10, Asn17, and Lys19, all of which are solvent-exposed. Thus, no major conformational changes occur in the monomer upon DNA binding or due to the engineered mutations. Details of protein−DNA interactions are shown in Figure 3 and discussed in the Discussion.

NMR Spectral Assignment of the CI-NTD. To map the site of interaction between the CI-NTD and the full palindromic DNA site in solution, we performed nuclear magnetic resonance (NMR) experiments. The complete assignment of the CI-NTD was obtained using a doubly labeled (^{13}C and ^{15}N) sample of the CI-NTD and a set of triple-resonance experiments (see Materials and Methods) (Figure 4). The secondary structure propensity (SSP)[40] calculated from the experimental Cα and Cβ chemical shifts shows that the secondary structure in solution is the same as that observed in the crystal structure (Figures 1A,B and 4). No chemical shift changes were observed when the protein concentration was changed, indicating that the CI-NTD remains a monomer at the highest concentration tested (7.8 mg/mL).

Interactions with DNA in Solution. Following the assignment of the ^{1}H−^{15}N HSQC spectrum, we conducted titrations of the CI-NTD with dsDNA corresponding to a full operator site. With an increase in the DNA concentration, some of the CI-NTD resonances shift according to a fast exchange regime, while some resonances disappear at substoichiometric amounts of DNA, in agreement with an intermediate exchange regime (Figure 5A). The disappearance of resonances or large changes in both chemical shifts and intensity are indicative of residues that interact with DNA. The beginning of helices α2 and α3 through to the beginning of helix α4 is clearly located in the interaction site (Figure 5B,C). Some intensity changes are also observed for some residues in α1, but the chemical shift differences are very small. Mapping of the interacting residues on the structure of the CI-NTD half-site presented here shows full agreement with the crystal structure (Figure 3A), although the interaction of α4 with a full palindromic site may in solution extend to the minor groove (see Discussion).

DISCUSSION

Monomeric State of the CI-NTD in Solution. Taken together, our results from gel filtration, SAXS, and NMR indicate that in the concentration range tested (up to ∼8 mg/mL), the CI-NTD is a monomer. In contrast, e.g., full-length CI forms a hexamer at concentrations as low as 0.7 mg/mL.[7] A construct with the last 58 C-terminal residues removed has been shown to be dimeric by gel filtration, suggesting that a dimerization region is present between residues 74 and 122 of CI.[7]

Structural Comparison. Structure similarity searches were conducted in DaliLite[41] to identify the homologues most structurally homologous to the CI-NTD. Levels of structural and sequence similarity to the first 20 DaliLite hits above a Z score threshold of 9.0 are summarized in Table S1 of the Supporting Information. The structures that arose from the DaliLite search were fairly consistent with the potential MR models initially identified by the PHYRE[18] sequence-based search, but surprisingly, the top scorer was not a known phage

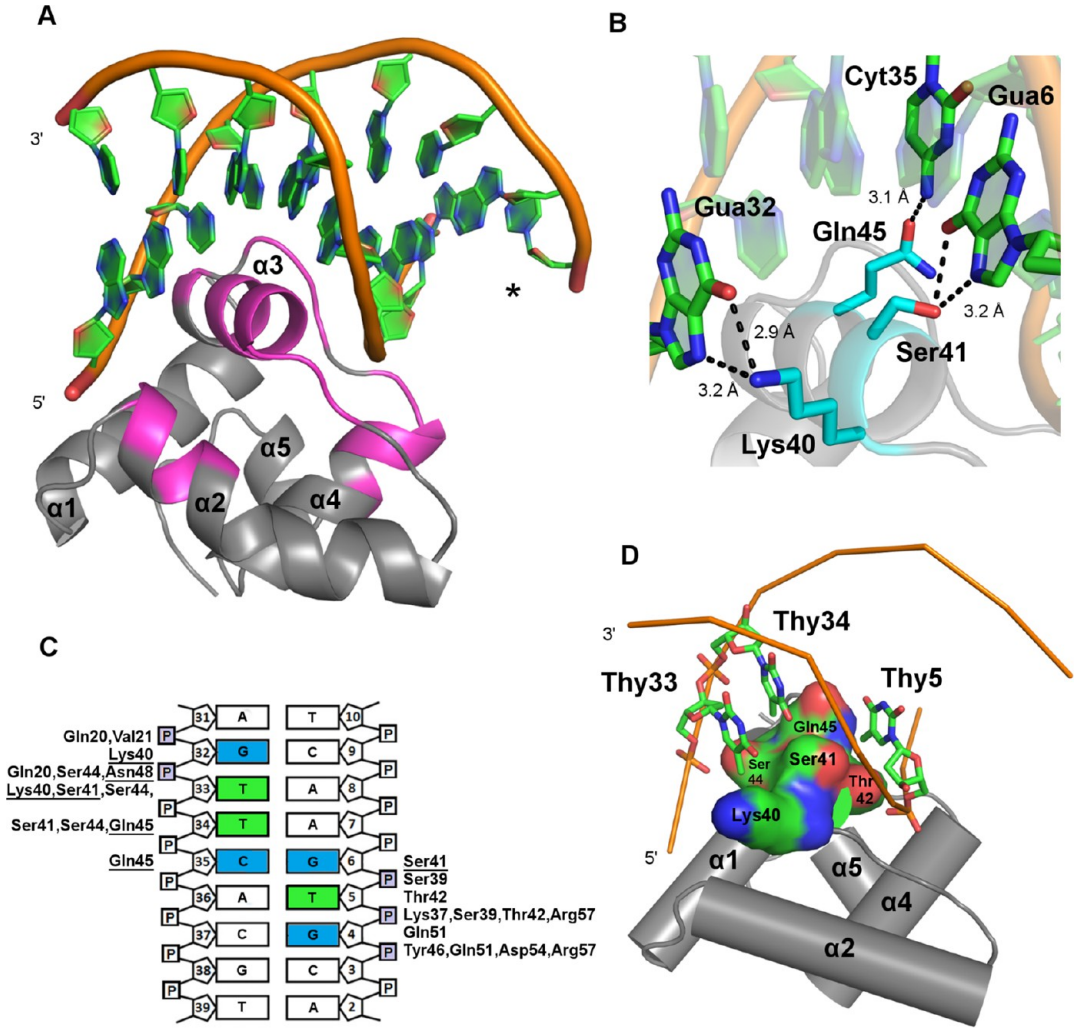


Figure 3. Interactions of the CI-NTD with DNA. (A) Overall view of interactions with DNA in the crystal structure of the complex of the CI-NTD with its half-operator site. The asterisk indicates where the center of the palindrome would be in a full operator. CI-NTD residues interacting with DNA in solution, as detected by NMR, are colored magenta. Specifically, residues whose resonances disappear upon interaction with DNA in the NMR titration (gray bars in Figure 5B,C) and residues experiencing both a significant change in chemical shift (above the dashed line in Figure 5B) and large intensity changes (below the dashed line in Figure 5C) are mapped. Gray regions are not perturbed by the presence of the DNA. (B) Close-up of the major groove and the CI-NTD recognition helix. The side chains involved in the most important base-specific interactions are shown as cyan sticks. (C) Schematic representation of interactions of the CI-NTD with DNA. Residues are aligned with the bases or phosphates that they contact. Contacted bases are highlighted in cyan and green for polar and apolar interactions, respectively. Contacted phosphates are highlighted in gray. Previously mutated residues shown to be important for DNA binding are underlined. (D) Surface representation of the recognition helix showing apolar contacts with selected DNA bases. Carbon atoms are mapped to the surface in green, oxygen atoms in red, and nitrogen atoms in blue.

protein, but the SinR protein of *Bacillus subtilis*, a key regulator for biofilm formation.[42,43] Even for this, only 22% sequence identity was observed, with an rmsd of 1.4 Å over 63 aligned residues. SinR is also part of a repressor/antirepressor system, but the molecular mechanisms are different. To the best of our knowledge, there are no MOR equivalent sequences in *Bacillus subtilis* and the molecular mechanism of antirepression involves heterodimer formation with SinI, which does not have a HTH motif, while full-length SinR forms a tetramer in the course of its function.

Further down the list some bacteriophage repressors appeared, including CI of phage 434[39] and the P22 C2[22] repressor NTD, which was used as the MR search model. Surprisingly, known CI repressors from other bacteriophages, including λ[2] and the 186 phage,[5] were not among the 50 first hits, but a comparison with these structures is also included in Table S1 of the Supporting Information.

DNA Binding. The NTDs of HTH transcription factors in the Cro and repressor family, for example, repressors from phage P22 and phage λ,[2,22,39] usually contact an operator site as a dimer, with each of the recognition helices contacting one major groove (see, e.g., structures of PDB entries 1RPE, 2R1J, and 1LMB). There are some exceptions; for example, it has been suggested for SinR that the NTDs binding adjacent half-sites come from two different dimers in the tetramer.[43] CI from TP901-1 is also believed to function as a trimer of dimers with two NTDs binding to a full operator site.[7] However, binding of engineered monomeric forms of HTH transcription factors to repressor half-sites, such as seen here for the complex of the CI-NTD with DNA, has previously been reported,[44] although with DNA binding affinity lower than that observed for dimeric

dx.doi.org/10.1021/bi400439y | *Biochemistry* 2013, 52, 6892−6904

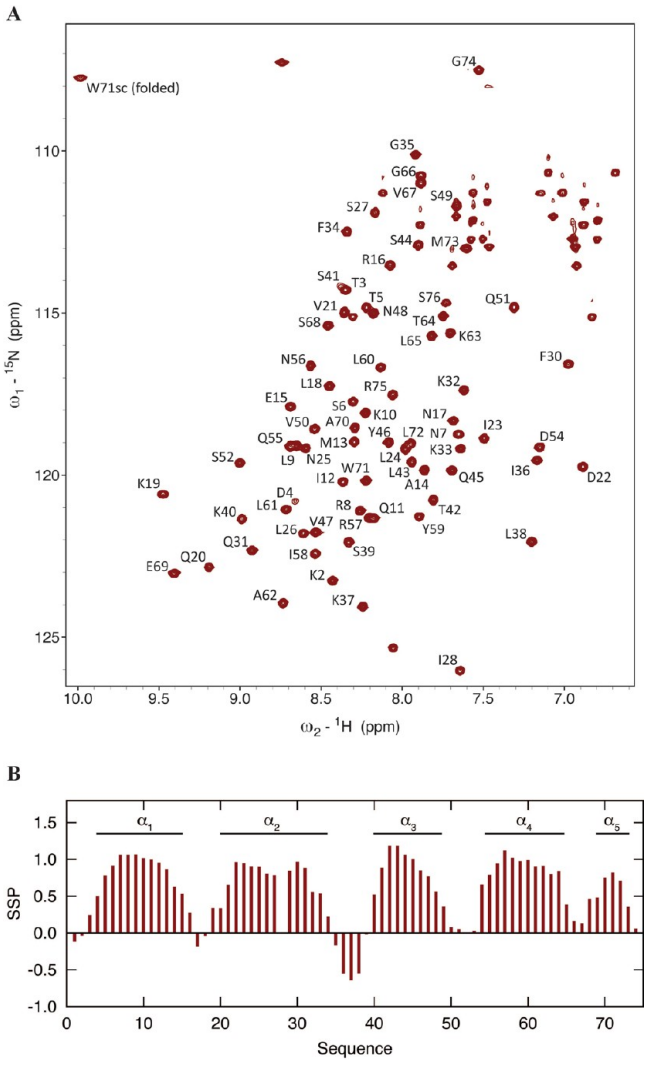


Figure 4. (A) Assignment of the ^{1}H−^{15}N HSQC spectrum of the NTD at 25 °C and pH 6.5 in 20 mM Tris buffer and 100 mM NaCl. (B) Secondary structure propensity (SSP) of the CI-NTD calculated from the experimental Cα and Cβ chemical shifts as described in ref 40. The horizontal black lines indicate the positions of the helical elements in the crystal structure of the CI-NTD.

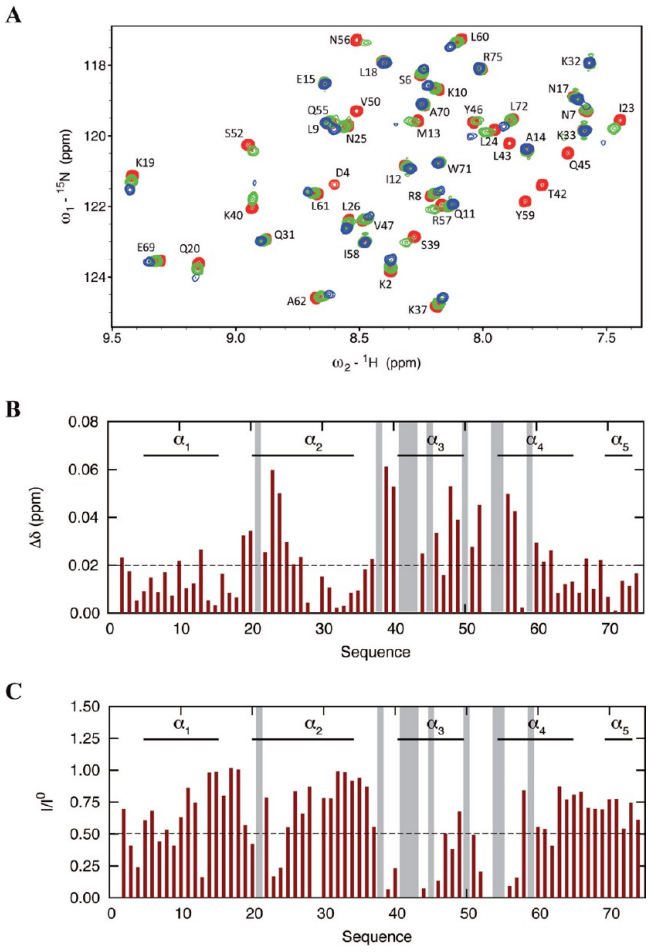


Figure 5. Mapping of the interaction site of the CI-NTD with dsDNA corresponding to a full operator site by NMR titration experiments. (A) Excerpt of ^{1}H−^{15}N TROSY spectra of the CI-NTD for different CI-NTD and DNA concentrations of 284 and 0 μM (red), 246 and 60 μM (green), and 166 and 192 μM (blue), respectively. (B) Combined chemical shift difference of ^{1}H and ^{15}N $\{\Delta\delta = [(\Delta\delta_N/6.5)^2 + (\Delta\delta_H)^2]^{1/2}\}$ between the first titration point without DNA and the titration point corresponding to CI-NTD and DNA concentrations of 231 and 84 μM, respectively. Gray bars indicate missing peaks in the spectrum due to exchange on the intermediate chemical shift time scale. The dashed line corresponds to the level above which the chemical shift changes were considered to be significant (chemical shift perturbations of >0.02 ppm). (C) Ratios of intensities (peak heights) in the TROSY spectra used in panel B. I^0 is the intensity in the spectrum of the isolated CI-NTD, while I is the intensity in the spectrum containing DNA. The dashed line corresponds to the level below which the changes in intensities were considered to be significant (ratio of <0.50).

versions. Because attempts to crystallize full-length and dimeric CI constructs have so far been unsuccessful, presumably because of their inherent flexibility, the CI-NTD complex with a half-site can be used to reveal the detailed interactions with DNA. We expect these interactions to be the same as those for full-length CI interacting with a full operator site, although information about, for example, the influence of binding on the full operator conformation is lost.

Binding of the CI-NTD to its half-operator site has similarities to that of previously reported complexes for the Cro and repressor HTH family, e.g., λ CI,[2] P22 C2,[22] and SinR[43] (Figure 6A,B), with a similar position of α3 (RH) in the major groove, where it recognizes specific bases (direct readout) partly dictating sequence specificity.

The first residue of the CI-NTD RH (Lys40) interacts with the second base pair of the DNA (Gua32 in the used numbering). Lys40 donates a hydrogen bond to both O6 and N7 of Gua32, explaining the discrimination against any other nucleobase at this position. The second residue (Ser41) and a residue one turn further downstream of the helix (Gln45) make contacts with base pair 5 (Gua6/Cyt35). Ser41 has possible hydrogen bonds to N7 and O6 of Gua6 through its hydroxyl, while the complementary base of Gua6, Cyt35, donates a hydrogen from N4 to the side chain amide oxygen of Gln45. Detailed interactions formed by these three residues are shown in Figure 3B. Gln51 also hydrogen bonds to N7 of Gua4 in base pair 7 through its side chain amide.

In addition to these hydrogen bonds, van der Waals interactions also seem to play a role in specificity. The aliphatic chains of Lys40 and Gln45 along with Cβ of Ser41 and Ser44 create a hydrophobic patch, which in the complex comes in

dx.doi.org/10.1021/bi400439y | *Biochemistry* 2013, 52, 6892−6904

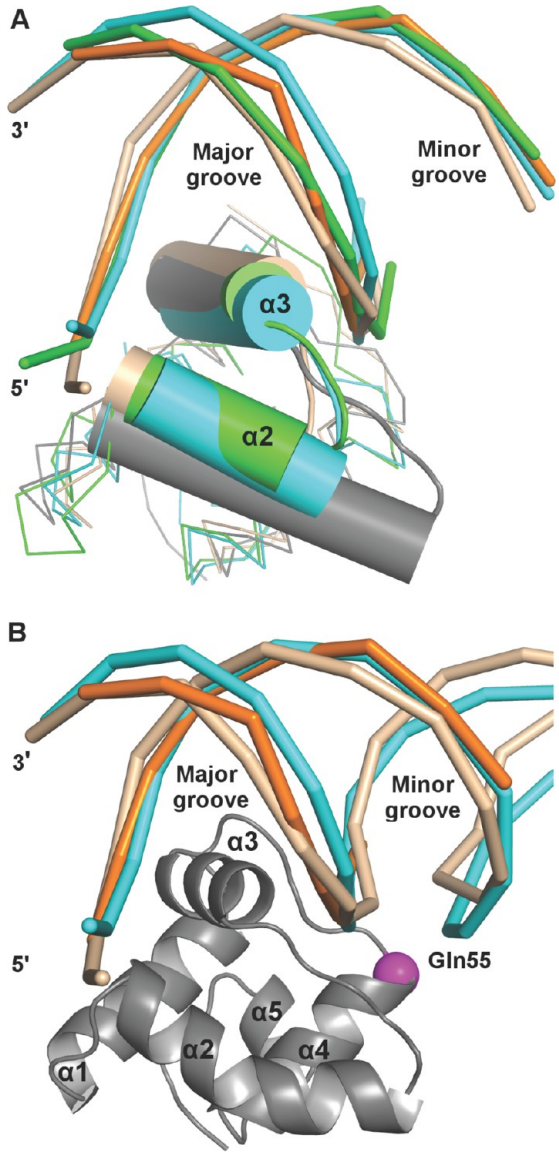


Figure 6. (A) Comparison of binding of phage repressor HTHs in the major groove. NTD−DNA complex structures of λ phage, P22 C2, and TP901-1 [green (PDB entry 1LMB), cyan (PDB entry 2R1J), and gray (PDB entry 3ZKC), respectively] are shown with helices in the HTH motifs represented as cylinders. (B) Comparison of DNA binding by SinR, P22 C2, and the CI-NTD. The operator DNA of SinR (PDB entry 3zkc) is colored pale yellow, the operator DNA of P22 C2 (PDB entry 2R1J) light blue, and the NTD of TP901-1 gray, and Gln55 Cα is represented by a magenta sphere. As supported by the NMR results, Gln55 may make minor groove interactions with a full operator site.

close contact with the methyl groups of Thy33 and Thy34 at the third and fourth positions, respectively (Figure 3D). A preference for Thy at the fourth position has been shown by biochemical experiments.[8] At the sixth position, the A:T motif positions another methyl close to the third residue of the RH, Thr42, potentially making favorable van der Waals interactions with Cα and Cγ.

The interactions of CI-NTD Lys40 with Thy33 (van der Waals) and Gua32 (hydrogen bonding), and of Gln51 with Gua4, are very similar to those made by SinR (Lys28 contacting Gua4 and Thy5 and Gln39 contacting Gua14).[43] The hydrophobic environment is to some extent reminiscent of the valine cleft of P22 C2.[22] Hydrophobic patches on the CI-NTD interacting with methyl groups of thymines in the major groove are comparable to those of SinR,[43] λ CI/Cro,[2,45] and phage 434 Cro.[46]

It has become clear that the sequence specificity of DNA-binding proteins is also dictated by sequence-specific conformations of the DNA double helix, resulting in indirect readout.[47] The indirect readout involves contacts to both the edges and the middle of the operator site. In this HTH family, contacts on the edge of the operator site frequently involve interactions with the first two to four phosphates formed by residues from α1, the amino-terminal part of the scaffolding helix (α2), and polar or charged residues from the carboxyl-terminal part of the recognition helix (RH/α3). The TP901-1 CI-NTD also makes interactions from α2 and α3 with at least two phosphates (e.g., through Asn48), but none from α1 are seen in the crystal structure (Figure 3C). The absence of interactions from α1 in the CI-NTD correlates with the absence of a salt bridge formed between an Arg in α1 and a Glu in α3, which is often found in phage repressors.[22,39] In the CI-NTD, hydrophobic residues (Met13−Val47) are present in equivalent positions. Previous studies of the Arg10Met variant of the CI repressor from phage 434 resulted in a translation of α1, which affected DNA binding, by altering the interaction network of Arg10 with polar residues in the second helix (α2).[48] Evidently, this salt bridge is not necessary for the function of TP901-1 CI.

In the middle of the operator site, the CI-NTD makes similar interactions with the sugar−phosphate backbone as also found in CI repressors from other phages,[2,22,39] through polar residues of loop 2, the side of α3 (RH), loop 3, and the N-terminal part of α4 (Figure 3A,C), but because in the case of the TP901-1 repressor we were able to crystallize only the NTD bound to a half-site, only limited conclusions can be drawn about the indirect readout.

The regions of the protein interacting with DNA in the crystal structure are in good agreement with those found from NMR in solution (Figure 3A). One exception is the beginning of helix α4, which shows clearer interactions with DNA in the NMR titration experiments than in the crystal structure. Asp54 and Arg57 are in the proximity of DNA in the crystal structure (see the diagram in Figure 3C), but for example, Gln55, which by NMR is shown to interact with the full-length palindromic operator, interacts in the crystal structure with a symmetry-related molecule. Because the crystal structure contains only a half-site, the NMR results suggest that with a full operator site this residue may contact a phosphate in the minor groove. A superposition with SinR and P22 C2 in complex with DNA (Figure 6B) shows that that with a full operator site Gln55 from the CI-NTD may contact a phosphate in the minor groove, consistent with the NMR results.

Previous mutational studies[7] are also consistent with the crystal structure. The double mutants K40A/S41A and Q45A/N48A were impaired in DNA binding, with apparent dissociation constants ~400-fold higher than that of the wild type. This is consistent with the involvement of the four residues in binding (Lys40, Ser41, and Gln45 in direct readout and Asn48 in indirect readout) as described above.

General Protein Sequence Rules for DNA Specificity in the Cro and Repressor HTH Family. As others have also observed previously, 1:1 sequence correspondence rules between proteins and bound DNA cannot be established, partly because there are significant effects from the indirect

dx.doi.org/10.1021/bi400439y | Biochemistry 2013, 52, 6892−6904

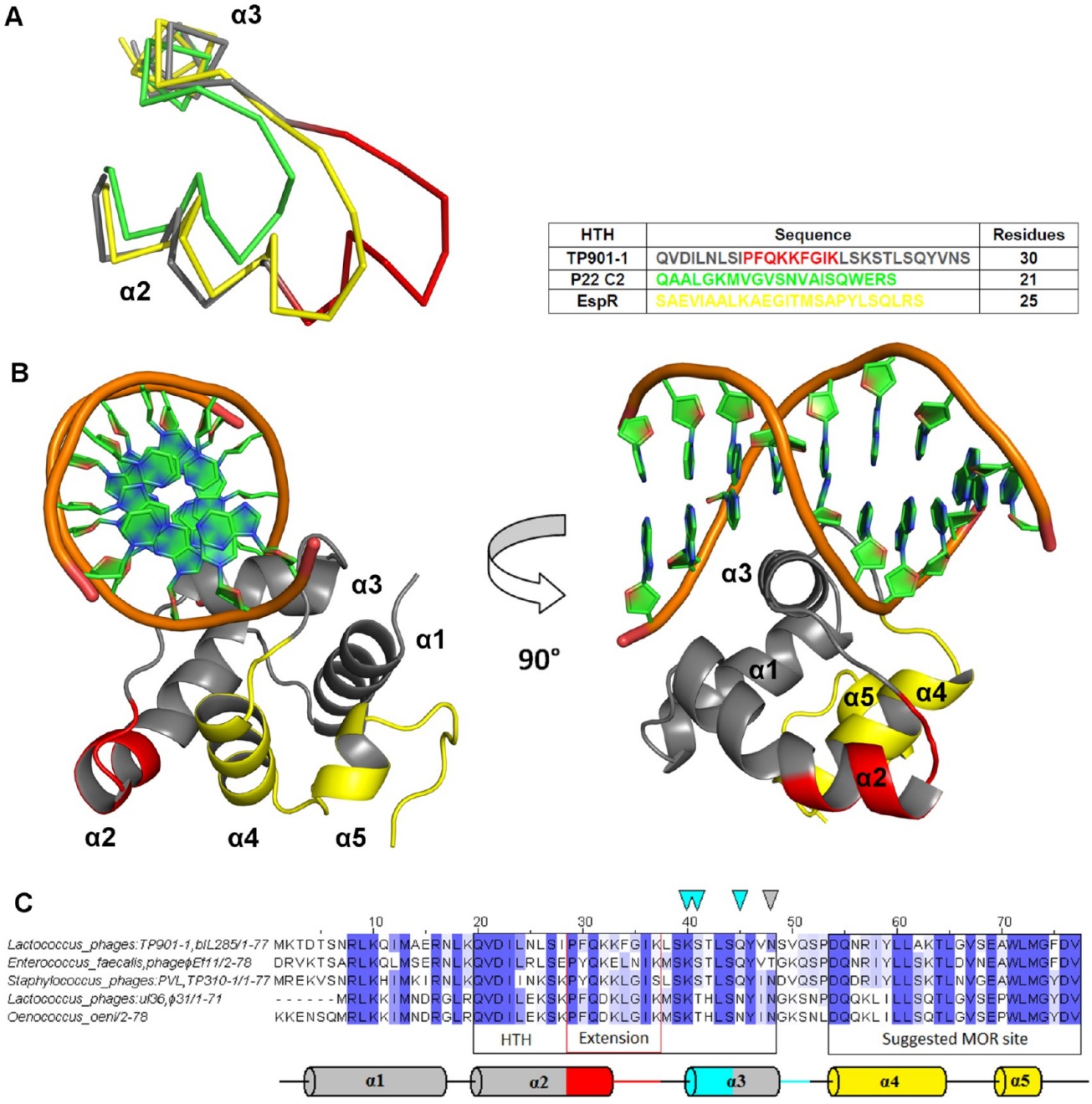


Figure 7. (A) HTH motifs of the CI-NTD (gray and red), P22 C2 (green, PDB entry 2R1J), and *M. tuberculosis* EspR (yellow, PDB entry 3QYX) superimposed and shown as ribbons. The HTH motif of P22 C2 comprises 21 residues and is a canonical HTH motif. In the HTH motif of TP901-1, the extended part of the scaffolding helix (nine additional residues relative to the classical HTH motif) is shown (red). The HTH motif of EspR also contains a longer scaffolding helix; however, it remains shorter than TP901-1. (B) CI-NTD (gray) with the proposed MOR interaction site[11] (yellow). The extension of the scaffolding helix in the HTH motif (red) is positioned close to the putative MOR site, further supporting the notion that the extended scaffolding helix plays a role in interactions between the CI-NTD and MOR. (C) Sequence alignment of the CI-NTD and related proteins. The HTH motif of TP901-1, the extension (red box), and the putative MOR interaction site are indicated. Mutations previously found to affect DNA binding are denoted with cyan and white triangles for those involved in direct and indirect readout, respectively. Secondary structure elements are aligned along the sequence and colored according to function.

readout; however, we observed some general trends. A Lys at the first position of the RH seems to give a preference for a GpT step at the first two recognized base pairs (as in TP901-1 and SinR), while a Gln in the first position of the RH corresponds to a preference for Ade at the second recognized base (as for λCI/Cro and 434CI/Cro). Residues at position 2 of the RH generally interact with the third and/or fourth recognized base pairs. Though there is here a preference for polar residues [Gln and Ser, except for P22 C2 (Val33)], the correlations with recognized bases are less clear. The third position of the RH is generally nonbulky, if the fifth recognized base is a Thy. For example, the CI-NTD has Thr42, while SinR, which recognizes an Ade instead of Thy, has Tyr30 at this position. Although other residues in the RH interact with DNA, no simple sequence rules can be extracted.

dx.doi.org/10.1021/bi400439y | *Biochemistry* 2013, 52, 6892−6904

Specificity for Different Operator Sequences. The observed interactions between the NTD and DNA in the crystal structure give a structural rationale for the observed difference in CI affinity for different operator sites.[8] CI was shown to have an almost 1000-fold lower affinity for O_R than for O_L but an affinity similar to that for O_D. While both O_L and O_D conform almost perfectly to the palindromic sequence (**AG**TT**C**AYGXXXCRT**G**AA**C**T), the O_{R1} half-site has an A:T pair at the second position rather than a G:C pair and an A:T motif at the eighth position instead of a G:C pair (A<u>A</u>TTC-AT<u>A</u>TT).

In Gua, the presence of a carbonyl group at C6 makes it a good hydrogen bond acceptor for lysine's amino group, whereas the corresponding amine in Ade is best as a hydrogen donor.[49] Thus, one of the two hydrogen bonds that Lys40 forms with Gua32 (the second base in the consensus) is not possible with Ade, partly explaining the decreased affinity for the O_{R1} half-site. Furthermore, it has already been shown by EMSA that substitution of the eighth G:C motif with A:T disfavors binding,[8] and thus, this substitution contributes also to the higher affinity of the CI-NTD for O_L and O_D than for O_R. The Gln51 side chain might be the residue that discriminates between a Cyt and a Thy by clashing with the thymine's methyl group.

The indirect readout will also have an influence on sequence specificity, but these effects are harder to predict on the basis of the presented structure.

Extended Scaffolding Helix in the CI-NTD and Possible Roles in Protein−Protein Interaction. As described above, the overall fold and mode of interaction with DNA of the CI-NTD are quite similar to those of other phage repressors, despite significant sequence dissimilarities. However, one distinct feature of the CI-NTD is an extension of the scaffolding helix (α2) and following loop (loop 2) (Figure 7A), resulting in a noncanonical HTH motif. Where the turn in other HTH motifs would usually begin, the scaffolding helix in the CI-NTD structure has a small kink at a Pro residue before an extension of the scaffolding helix and of the loop connecting it to the RH, from which point on the structure of the CI-NTD superimposes well again onto other repressors. The extension has a PFQKKFGIK sequence and has minimal involvement in DNA binding. The transcriptional regulator EspR of *Mycobacterium tuberculosis* that activates the ESX-1 secretion system[36] is the only one from the first 20 DaliLite hits (see Table S1 of the Supporting Information) that shows an extended scaffolding helix relative to the classical HTH motif, and even in this case not to the same extent as for the CI-NTD (Figure 7A). The EspR repressor is unusual because it is incapable of binding to consecutive major grooves.[36] Therefore, it has been proposed that the EspR dimer binds with one subunit to the major groove of the operator site, while the other subunit is turned away to potentially bind at a more distant operator site. In the dimeric structure of EspR (PDB entry 3QYX), the two scaffolding helices are part of the monomer−monomer interface.

The palindromic operator sites in the genetic switch of TP901-1 point to a canonical binding of a TP901-1 CI dimer to adjacent major grooves. However, the extension of the scaffolding helix in the CI-NTD could also play a role in protein−protein interactions as in EspR, perhaps by participating in an interaction with the MOR antirepressor.

A CI/MOR interaction site in the CI-NTD was previously suggested in the latter part of the sequence because of a high degree of sequence similarity between CI from organisms also containing MOR homologues[11] and is positioned close to the extension of the scaffolding helix (Figure 7B).

Related Systems. The CI/MOR (repressor/antirepressor system) in TP901-1 is almost identical at the sequence level to that found in the lactococcal prophage bIL285 (Figure 7C). The operator sites are also identical, so the switch is expected to function in the same way as in TP901-1. CI/MOR homologues are found in several virulent lactococcal phages belonging to the P335 group, which is problematic in the dairy industry, for example, ul36 and Φ31 (Figure 7C). The CI sequence of another virulent phage, P335, is almost identical to those of ul36 and Φ31,[50−52] and all the sequences are ∼60% identical with that of TP901-1 CI, including conservation of the extended helix. Previous work on Φ31 has shown that while the wild-type repressor from this phage cannot bind to the operator sites *in vivo*, a cloned C-terminally truncated CI repressor is able to bind specific operator sites and hence repress both Φ31 and related virulent phages when the bacteria are infected.[53] Using the truncated repressor, the operator sites were identified both *in vivo* and *in vitro*. Thus, we can conclude that this group of virulent phages contains functional operator sites and that the NTD of the repressor has retained the ability to repress when overexpressed. It might be suggested that the inability of the wild-type CI repressor in these phages to repress is due to the presence of an unidentified antirepressor. However, Φ31 also cannot establish lysogeny because it lacks the necessary phage attachment site and integrase. The CIs from these virulent lactococcal phages are also very highly related (>95% identical sequence) to putative proteins from *Oenococcus oeni*, a Gram-positive lactic acid bacterium used in wine making, which also possesses a MOR homologue.

Interestingly, BLAST searches with the CI-NTD sequence find several putative proteins from Gram-positive bacterial pathogens (e.g., Streptococci, Staphylococci, and Enterococci) and their phages, about which very little structural or functional information is known. In fact, when considering only NTDs, TP901-1 seems to have a closer relation to phages of *Staphylococcus* and *Enterococcus* (such as ΦPVL and ΦEf11) than the virulent *Lactococcus* phages (P335, ul36, and Φ31). The sequences are very similar in the NTDs (maximal level of identity of approximately 50−75%) and somewhat similar in the dimerization region (residues 74−122), while the latter part of the CTD remains rather different. Two regions in the CI-NTD, residues 18−24 (LKQVDI) and residues 39−48 (SKSTLSQYVN), involved in DNA recognition are particularly conserved. Occasionally, very conservative substitutions occur for nucleobase-contacting residues, though with identical or comparable functional groups and properties (Ser41 to Thr, Gln45 to Asn, and Asn48 to Thr) (Figure 7C). Furthermore, Asp54 and Gln55 are highly conserved possibly to stabilize a narrow minor groove upon binding. In the scaffolding helix, an S-P-Q pattern (residues 27, 29, and 31, respectively, in TP901-1) is very highly conserved. This pattern (from the Pro) marks the start of the helix extension. In the crystal structure, Pro29 kinks the helix while Ser27 makes a hydrogen bond with Gln31. The high degree of sequence conservation in the helix extension (Figure 7C) testifies to its importance. Furthermore, the hydrophobic core of the NTD is also rather conserved, which further highlights the necessity of a conserved NTD scaffold. A BLAST search performed with the TP901-1 MOR sequence reveals highly conserved MOR homologues (approximately 50−65% identical sequence) for all the bacterial

dx.doi.org/10.1021/bi400439y | *Biochemistry* 2013, 52, 6892−6904

pathogens (or their phages) containing a CI homologue. It must be stressed that all these organisms appear to contain an extended helix in their putative repressors, further supporting a functional role of the helix extension, which could be interaction with the MOR homologues.

The role of phages and phage genome remnants in the virulence of pathogenic bacteria is more frequently being recognized.[54] The similarity of the TP901-1 CI repressor to temperate staphylococcus phage ΦPVL suggests that mechanistic knowledge of the genetic switch may advance the understanding of pathogenesis of methicillin-resistant *Staphylococcus aureus* (MRSA) strains,[55,56] where ΦPVL genes have been suggested as virulence factors, and help in antibiotic design.

CONCLUSION

The HTH motif is a widespread scaffold for DNA binding, with many features conserved in the vast majority of systems, yet it participates in molecular mechanisms with surprisingly diversified possibilities. In the genetic switches of temperate phages, this diversity is partly achieved by being coupled with different linkers and C-terminal oligomerization domains. In the CI-NTD, we observe an extension of the scaffolding helix, an element we suggest could be involved in protein−protein interactions. To the best of our knowledge, this is the most extended version of the helix in HTH motifs, and the sequence in this region and for the CI-NTD overall is very closely related to structurally and functionally uncharacterized proteins from Gram-positive pathogenic bacteria and their phages. The putative conserved fold (including the placement of the extended scaffolding) indicates repressor−antirepressor interactions may also be important in these organisms. Understanding the TP901-1 genetic switch and the MOR−CI (NTD) interplay and hence the design of antirepressors not only might provide beneficial insight into beverage and food processing but also might assist in the treatment of the otherwise antibiotic multiresistant (such as MRSA) strains now frequently emerging.

ASSOCIATED CONTENT

Supporting Information

Supplementary table that includes details about the structural and sequence similarity to the TP901-1 CI-NTD for several structural homologues. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

Coordinates and structure factors for the CI-NTD and its complex with DNA have been deposited in the Protein Data Bank as entries 3zhi and 3zhm, respectively. NMR assignments have been deposited in the BioMagResBank (BMRB) as entry 19326.

AUTHOR INFORMATION

Corresponding Author

*Department of Chemistry, Universitetsparken 5, 2100 Copenhagen Ø, Denmark. Telephone: +45 35320295. Fax: +45 35320322. E-mail: leila@chem.ku.dk.

Present Address

[@]M.P.: Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Fremtidsvej 3, Hørsholm, Denmark.

Author Contributions

K.H.F. and K.K.R. contributed equally to this work.

Funding

This work was supported by the Danish Council for Independent Research (Grant 10-084503) and the French Agence National de la Recherche through ANR JCJC ProteinDisorder (to M.R.J.). A Ph.D. studentship to K.K.R. was jointly funded by the University of Copenhagen and the Danish Council for Independent Research (Grant 09-065434). Travel to synchrotrons was supported by the DANSCATT program, funded by the Danish Council for Independent Research (09-074479) and the ELISA program, funded by the European Community's Seventh Framework Programme (226716).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank F. Fredslund and Martin Blackledge for helpful discussions on crystallographic programs and NMR data, respectively, D. Boelskifte for technical assistance, and the synchrotrons (MAXLAB and ESRF) for beamtime. Furthermore, we thank Prof. Jesper Bendix for the overall administration of the cofinanced Ph.D. scholarships.

ABBREVIATIONS

CTD, C-terminal domain; CI-NTD, wild-type N-terminal domain of the TP901-1 CI repressor; CI-NTDK40A/S41A, variant of the CI-NTD with residues 40 and 41 mutated to Ala; CI-NTDQ45A/N48A, variant of the CI-NTD with residues 45 and 48 mutated to Ala; EMSA, electrophoretic mobility shift assay; HTH, helix−turn−helix; LLG, log-likelihood gain; MR, molecular replacement; NTD, N-terminal domain; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; RH, recognition helix; RFZ, rotation function *Z* score; rmsd, root-mean-square deviation; SAXS, small angle x-ray scattering; TFZ, translation function *Z* score.

REFERENCES

(1) Brøndsted, L., and Hammer, K. (2006) Phages of *Lactococcus lactis*. In *The Bacteriophages* (Calendar, R., and Abedon, S. T., Eds.) 2nd ed., pp 572−592, Oxford University Press, New York.

(2) Beamer, L. J., and Pabo, C. O. (1992) Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol. 227*, 177−196.

(3) Ptashne, M. (2004) *A Genetic Switch: Phage Lambda Revisited*, 3rd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.

(4) Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., and Lewis, M. (2008) Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. *Nature 452*, 1022−1025.

(5) Pinkett, H. W., Shearwin, K. E., Stayrook, S., Dodd, I. B., Burr, T., Hochschild, A., Egan, J. B., and Lewis, M. (2006) The structural basis of cooperative regulation at an alternate genetic switch. *Mol. Cell 21*, 605−615.

(6) Shearwin, K. E., Brumby, A. M., and Egan, J. B. (1998) The Tum protein of coliphage 186 is an antirepressor. *J. Biol. Chem. 273*, 5708−5715.

(7) Pedersen, M., Lo Leggio, L., Grossmann, J. G., Larsen, S., and Hammer, K. (2008) Identification of quaternary structure and functional domains of the CI repressor from bacteriophage TP901-1. *J. Mol. Biol. 376*, 983−996.

(8) Johansen, A. H., Brondsted, L., and Hammer, K. (2003) Identification of operator sites of the CI repressor of phage TP901-1: Evolutionary link to other phages. *Virology 311*, 144−156.

(9) Alsing, A., Pedersen, M., Sneppen, K., and Hammer, K. (2011) Key Players in the Genetic Switch of Bacteriophage TP901-1. *Biophys. J. 100*, 313−321.
(10) Madsen, P. L., Johansen, A. H., Hammer, K., and Brondsted, L. (1999) The genetic switch regulating activity of early promoters of the temperate lactococcal bacteriophage TP901-1. *J. Bacteriol. 181*, 7430−7438.
(11) Pedersen, M., Ligowska, M., and Hammer, K. (2010) Characterization of the CI Repressor Protein Encoded by the Temperate Lactococcal Phage TP901-1. *J. Bacteriol. 192*, 2102−2110.
(12) Kaneko, J., Kimura, T., Narita, S., Tomita, T., and Kamio, Y. (1998) Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage ϕ PVL carrying Panton-Valentine leukocidin genes. *Gene 215*, 57−67.
(13) Stevens, R. H., Ektefaie, M. R., and Fouts, D. E. (2011) The annotated complete DNA sequence of *Enterococcus faecalis* bacteriophage Ef11 and its comparison with all available phage and predicted prophage genomes. *FEMS Microbiol. Lett. 317*, 9−26.
(14) Stanley, E., Fitzgerald, G. F., LeMarrec, C., Fayard, B., and vanSinderen, D. (1997) Sequence analysis and characterization of ϕ O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology 143*, 3417−3429.
(15) Holden, M. T. G., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P. J., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S. D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K. D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B. G., Spratt, B. G., and Parkhill, J. (2004) Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. U.S.A. 101*, 9786−9791.
(16) Kabsch, W. (2010) Xds. *Acta Crystallogr. D66*, 125−132.
(17) Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D67*, 271−281.
(18) Kelley, L. A., and Sternberg, M. J. E. (2009) Protein structure prediction on the Web: A case study using the Phyre server. *Nat. Protoc. 4*, 363−371.
(19) McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D63*, 32−41.
(20) Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658−674.
(21) Bunkoczi, G., and Read, R. J. (2011) Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D67*, 303−312.
(22) Watkins, D., Hsiao, C. L., Woods, K. K., Koudelka, G. B., and Williams, L. D. (2008) P22 c2 repressor-operator complex: Mechanisms of direct and indirect readout. *Biochemistry 47*, 2325−2338.
(23) Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D64*, 61−69.
(24) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126−2132.
(25) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D66*, 486−501.
(26) Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D68*, 352−367.
(27) Mylonas, E., and Svergun, D. I. (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr. 40*, S245−S249.
(28) Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr. 36*, 1277−1282.
(29) Svergun, D. I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr. 25*, 495−503.
(30) Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOL: A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr. 28*, 768−773.
(31) Lescop, E., Schanda, P., and Brutscher, B. (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson. 187*, 163−169.
(32) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR 6*, 277−293.
(33) Goddard, T. D., and Kneller, D. G. (2008) *SPARKY 3*, University of California, San Francisco.
(34) Jung, Y. S., and Zweckstetter, M. (200) Mars; Robust automatic backbone assignment of proteins. *J. Biomol. NMR 30*, 11−23.
(35) Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A. 94*, 12366−12371.
(36) Rosenberg, O. S., Dovey, C., Tempesta, M., Robbins, R. A., Finer-Moore, J. S., Stroud, R. M., and Cox, J. S. (2011) EspR, a key regulator of *Mycobacterium tuberculosis* virulence, adopts a unique dimeric structure among helix-turn-helix proteins. *Proc. Natl. Acad. Sci. U.S.A. 108*, 13450−13455.
(37) Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol. 1*, REVIEWS001.
(38) Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., and Iyer, L. M. (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev. 29*, 231−262.
(39) Shimon, L. J. W., and Harrison, S. C. (1993) The Phage 434 O(R)2/R1−69 Complex at 2.5-Angstrom Resolution. *J. Mol. Biol. 232*, 826−838.
(40) Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci. 15*, 2795−2804.
(41) Holm, L., and Rosenstrom, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res. 38*, W545−W549.
(42) Colledge, V. L., Fogg, M. J., Levdikov, V. M., Leech, A., Dodson, E. J., and Wilkinson, A. J. (2011) Structure and Organisation of SinR, the Master Regulator of Biofilm Formation in *Bacillus subtilis*. *J. Mol. Biol. 411*, 597−613.
(43) Newman, J. A., Rodrigues, C., and Lewis, R. J. (2013) Molecular basis of the activity of SinR protein, the master regulator of biofilm formation in *Bacillus subtilis*. *J. Biol. Chem. 288*, 10766−10778.
(44) Sauer, R. T., Pabo, C. O., Meyer, B. J., Ptashne, M., and Backman, K. C. (1979) Regulatory Functions of the Lambda-Repressor Reside in the Amino-Terminal Domain. *Nature 279*, 396−400.
(45) Albright, R. A., and Matthews, B. W. (1998) How Cro and lambda-repressor distinguish between operators: The structural basis underlying a genetic switch. *Proc. Natl. Acad. Sci. U.S.A. 95*, 3431−3436.
(46) Mondragon, A., and Harrison, S. C. (1991) The Phage-434 Cro/Or1 Complex at 2.5 Å Resolution. *J. Mol. Biol. 219*, 321−334.
(47) Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem. 79*, 233−269.

(48) Pervushin, K., Billeter, M., Siegal, G., and Wuthrich, K. (1996) Structural role of a buried salt bridge in the 434 repressor DNA-binding domain. *J. Mol. Biol. 264*, 1002−1012.
(49) Neidle, S. (2008) Principles of Protein-DNA Recognition. In *Principles of Nucleic Acid Structure*, pp 249−282, Academic Press, New York.
(50) Labrie, S., and Moineau, S. (2002) Complete genomic sequence of bacteriophage ul36: Demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. *Virology 296*, 308−320.
(51) Labrie, S. J., Josephsen, J., Neve, H., Vogensen, F. K., and Moineau, S. (2008) Morphology, genome sequence, and structural proteome of type phage P335 from *Lactococcus lactis*. *Appl. Environ. Microbiol. 74*, 4636−4644.
(52) Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D., and Chopin, M. (2001) Analysis of six prophages in *Lactococcus lactis* IL1403: Different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res. 29*, 644−651.
(53) Durmaz, E., Madsen, S. M., Israelsen, H., and Klaenhammer, T. R. (2002) *Lactococcus lactis* lytic bacteriophages of the P335 group are inhibited by overexpression of a truncated CI repressor. *J. Bacteriol. 184*, 6532−6544.
(54) Gal-Mor, O., and Finlay, B. B. (2006) Pathogenicity islands: A molecular toolbox for bacterial virulence. *Cell. Microbiol. 8*, 1707−1719.
(55) Lo, W. T., and Wang, C. C. (2011) Panton-Valentine Leukocidin in the Pathogenesis of Community-associated Methicillin-resistant *Staphylococcus aureus* Infection. *Pediatric Neonatology 52*, 59−65.
(56) Gordon, R. J., and Lowy, F. D. (2008) Pathogenesis of methicillin-resistant *Staphylococcus aureus* infection. *Clin. Infect. Dis. 46*, S350−S359.