*Published online 19 November 2013* *Nucleic Acids Research, 2014, Vol. 42, No. 4* **2725–2735**
*doi:10.1093/nar/gkt1119*

# Structural insight into DNA binding and oligomerization of the multifunctional Cox protein of bacteriophage P2

**Ronnie P.-A. Berntsson[1], Richard Odegrip[2], Wilhelmina Sehlén[2], Karin Skaar[1], Linda M. Svensson[1], Tariq Massad[1], Martin Högbom[1], Elisabeth Haggård-Ljungquist[2] and Pål Stenmark[1,*]**

[1]Department of Biochemistry and Biophysics, Arrhenius Laboratories for Natural Sciences, Stockholm University, SE-10691 Stockholm, Sweden and [2]Department of Molecular Biosciences, The Wenner-Gren Institute, Arrhenius Laboratories for Natural Sciences, Stockholm University, SE-10691 Stockholm, Sweden

Received July 11, 2013; Revised October 22, 2013; Accepted October 23, 2013

## ABSTRACT

**The Cox protein from bacteriophage P2 is a small multifunctional DNA-binding protein. It is involved in site-specific recombination leading to P2 prophage excision and functions as a transcriptional repressor of the P2 Pc promoter. Furthermore, it transcriptionally activates the unrelated, defective prophage P4 that depends on phage P2 late gene products for lytic growth. In this article, we have investigated the structural determinants to understand how P2 Cox performs these different functions. We have solved the structure of P2 Cox to 2.4 Å resolution. Interestingly, P2 Cox crystallized in a continuous oligomeric spiral with its DNA-binding helix and wing positioned outwards. The extended C-terminal part of P2 Cox is largely responsible for the oligomerization in the structure. The spacing between the repeating DNA-binding elements along the helical P2 Cox filament is consistent with DNA binding along the filament. Functional analyses of alanine mutants in P2 Cox argue for the importance of key residues for protein function. We here present the first structure from the Cox protein family and, together with previous biochemical observations, propose that P2 Cox achieves its various functions by specific binding of DNA while wrapping the DNA around its helical oligomer.**

## INTRODUCTION

Bacteriophage P2 is a non-lambdoid phage that can infect several enteric bacterial species (1). It is the prototype of the P2-like family of phages found in gamma-proteobacteria, and in natural *Escherichia coli* isolates almost 30% contain a P2-like prophage (2). After infection, P2 can either form lysogeny or grow lytically, which ultimately leads to cell lysis and the release of progeny phage particles. In the former case, the infected cells survive and the phage's DNA gets integrated into the chromosome of its host organism. The protein P2 Cox (control of excision), a small 91 amino acids long multifunctional protein, has previously been shown to have an intrinsic role in these processes. P2 Cox plays a role in P2 prophage excision, transcriptional repression of the P2 Pc promoter and in transcriptional activation of the satellite phage P4 $P_{LL}$ promoter. Phage P4 is dependent on structural genes and lysis functions of phage P2 for lytic growth [for a review, see (3)].

The lysogenic state of temperate phages is controlled by a regulatory protein, the immunity repressor. In phage P2, this task is performed by the C protein, which binds to the operator region of the P2 early operon containing the *cox* gene and thereby repressing the expression of this operon. The C protein is transcribed in the opposite direction to the early operon and contains a helix-turn-helix (HTH) motif that binds direct repeats in the operator region (4). The region between these coding regions is ~110 base pairs and contains the converging promoters of the early operon Pe, and the repressor Pc (5–7). The product of the

*To whom correspondence should be addressed. Tel: +46 8 163729; Fax: +46 8 153679; Email: stenmark@dbb.su.se
Present address: Tariq Massad, Therapure Biopharma Inc., 2585 Meadowpine Blvd, Mississauga, Canada.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

*cox* gene on the other hand inhibits the synthesis of the immunity repressor C. By binding to the Pc promoter, P2 Cox acts as a repressor of the lysogenic operon of the phage, controlling the expression of the immunity repressor and the integrase (8). In this aspect, it is thus functionally equivalent to the Cro protein of phage λ (9).

Early findings showed that P2 *cox* defective prophages were unable to produce free phages spontaneously during growth (10). In site-specific recombination, P2 Cox functions as a directionality factor by preventing integrative recombination and promoting excisive recombination (11). In this sense it is functionally equivalent to bacteriophage λ exisionase ($^{\lambda}$Xis) (12–14).

The satellite phage P4 carries the genes for transcriptional control, DNA replication and lysogenization, but it cannot perform lytic growth without structural genes and lysis machinery from a helper, e.g. the P2 phage (15). The derepression of phage P4 lysogens requires an intact P2 *cox* gene, and the binding site for P2 Cox has been located to a region upstream of the P4 late promoter, $P_{LL}$ (16).

The native form of the P2 Cox protein in solution is multimeric (17), and upon binding to its target it protects a region of at least 70 nucleotides from DNAse I cleavage (11,16). A comparison of the protected regions shows a consensus sequence TTAAAG/CNCA, denoted cox-boxes, which are present in at least six copies in the binding sites. Interestingly, the direction of the cox-boxes varies within the different targets and binding induces a strong bend in the DNA (18).

The crystal structures of the λ Cro and Xis proteins have been determined (9,19). The 66 amino acids long $^{\lambda}$Cro protein is a typical example of a HTH DNA-binding protein that binds to its 17 bases long operator sequence as a dimer and bends the DNA (12,13,20,21). The 77 amino acids long $^{\lambda}$Xis protein forms a winged HTH motif with an unstructured C-terminal tail (19,22). The crystal structure of a truncated version of $^{\lambda}$Xis, termed $^{\lambda\Delta 55}$Xis, lacking the last 27 residues, has been determined together with its 33 nucleotides long DNA target, indicating that $^{\lambda}$Xis forms a nucleoprotein filament with its DNA target (20,23). There is no or very low sequence similarities between the P2 Cox and λ Cro (9%) or Xis (20%) proteins, but a secondary prediction indicates that P2 Cox belongs to the winged HTH family of DNA-binding proteins (18,24,14).

In this article, we have determined the structure of full-length P2 Cox protein, which constitutes the first structure from the Cox protein family. The structure reveals that P2 Cox indeed has a winged HTH structure and, despite low sequence similarities, is structurally closely related to $^{\lambda}$Xis. Interestingly, the structure suggests that P2 Cox oligomerizes as a helical filament, and in that state binds and winds up DNA around the helical filament. Genetic and functional studies indicate the importance of the functional residues, and we propose that the C-terminal domain of P2 Cox is required for its oligomerization and function. Our study provides insight into the structure and function of the Cox proteins and the unexpected properties of the C-terminal region of these proteins.

## MATERIALS AND METHODS

### Site-directed mutagenesis

All constructs for the alanine screening were performed using the QuikChange™ Site-Directed Mutagenesis Kit or the QuikChange™ Lightning Multi Site-Directed Mutagenesis Kit, Stratagene, USA. The primers used for mutagenesis were obtained from either Eurofins MWG Operon, Germany, Thermo Fischer Scientific, Germany or DNA technology A/S, Denmark. All constructs were verified by DNA sequencing by Eurofins MWG Operon and Macrogen Inc., Korea. The primers used are listed in Supplementary Table S1.

### Protein purification

P2 Cox protein was purified from *E. coli* BL21(DE3) pLysE (25) containing plasmid pEE720 (17). The bacteria were grown in M9 minimal medium in a LEX bubbling system (SGC, Toronto, Canada) at 37°C until $OD_{600}$ 0.6 when 30 ml/l of amino acid mixture [100 mg each of lysine, threonine, phenylalanine, 50 mg each of leucine, isolecine, valine and L(+)-selenomethionine] was added. After another 20 min, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. The culture was allowed to grow overnight at 22°C until harvested by centrifugation. The pellet was resuspended in Buffer C (0.3 M potassium phosphate buffer, pH 7.5, 3 mM EDTA, 0.5 M KCl, 0.1% Triton, 0.5 mM TCEP) at 6 ml per gram of cells and freeze thawed twice to allow leakage of lysozyme and partial lysis. To complete lysis, resuspended cells were sonicated on ice in four 30 s bursts at 12–14 μm with an MSE Soniprep 150. The extract was clarified by centrifugation in a Sorvall RC5C at 23 000*g* for 1 h at 4°C, and ammonium sulfate was added to 25% saturation to the supernatant. After being gently stirred at 4°C for 1 h, the mixture was centrifuged at 17 000*g* for 30 min, and the pellet was subsequently resuspended in Buffer I (20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 15% isopropanol, 0.5 mM TCEP) followed by filtration through 0.45 and 0.22 μm filters. The extract was loaded on a HiLoad 16/60 Superdex200 column (GE Healthcare, Sweden). The Cox-containing fractions were analyzed by 20% homogenous SDS–PAGE gels using the PhastSystem (GE Healthcare) and concentrated using Vivaspin Centrifugal Concentrator (Vivaproducts, MA, USA). When purifying the non-labeled Cox protein an additional anion-exchange purification step was performed after the ammonium sulfate precipitation using HiTrap DEAE columns (GE Healthcare). Proteins were eluted at 2 M KCl followed by desalting on PD-10 columns (GE Healthcare) pre-equilibrated with Buffer II (20 mM Tris–HCl, pH 7.5, 150 mM NaCl) before gel filtration. Both purification procedures resulted in a very low protein yield (<1%), probably reflecting the tendency of P2 Cox to form higher order complexes in solution, which were subsequently lost in the void volume on the size exclusion chromatography column.

### Crystallization and structure determination

Crystallization was performed at 5.5 mg/ml of Cox using sitting drop vapor diffusion, using a reservoir solution containing 0.2 M Na-citrate pH 5.0, 24% PEG 4000 and 20% isopropanol, in a 1:1 drop ratio. Diffraction quality crystals grew at 18°C and appeared after 4–12 weeks. Crystals were frozen in liquid nitrogen without any further cryo-protection. Native diffraction data were collected at beamline ID14-2 at ESRF, France. The structure was solved using single-wavelength anomalous diffraction (SAD) collected at beamline PX1 at the SLS, Switzerland. Data processing and reduction were performed using XDS and programs from the CCP4 suite (26,27). Relevant statistics for phasing and refinement can be found in Table 1. Weak initial phase information was found using Phenix AutoSOL (28), which gave an initial model of 43 residues. Only one of the five Selenium atoms per monomer was located, with an σ level of 4. Numerous rounds of manual model building in Coot (29), and subsequent refinement in Refmac5 and Phenix.refine (28,30), were needed to complete the model. The native dataset diffracted to 2.4 Å, although had sub-optimal data completeness (92.9% overall). Unfortunately it was not isomorphous to the complete 2.9 Å SeMet dataset. The final model was therefore compared to the complete SeMet dataset. No differences were found, and since the electron density of the high resolution dataset was substantially better that one was used for the final round of model refinement. The final model was verified using Molprobity (31), and got a Molprobity score residing in the 97th percentile. All residues were within the Ramachandran allowed regions. The $R_{work}$ and $R_{free}$ of the final model were 20.9% and 25.3%, respectively. All structure figures were prepared using PyMOL.

### Complementation assay

Strain C-6005 was transformed with pEE720 derivatives containing the wild-type *cox* gene (17) or various alanine substitutions. Strain C-6005 (8) is a C-1 a derivative [$F^-$ prototrophic *E. coli* C strain (32)], lysogenized by the *cox* defective P2 mutant *cox3*. The transformants were grown at 30°C overnight in LB, supplemented with ampicillin 100 μg/ml for the transformed cells and with addition of 0.2 M potassium phosphate buffer pH 6.8 to avoid re-adsorption of released phages. The bacterial (CFU/ml) and phage titers (PFU/ml) were determined after dilution in triplicates. C-1757 [polyauxotrophic *E. coli* C strain *strp supD* (33)] was used as an indicator strain for phage titrations.

**Table 1.** Data collection and refinement statistics

| | Native | SeMet |
|---|---|---|
| Data collection | | |
| Space group | $P6_5$ | $P6_5$ |
| Cell dimensions | | |
| *a*, *b*, *c* (Å) | 66.3, 66.3, 33.8 | 64.8, 64.8, 33.9 |
| α, β, γ (°) | 90.0, 90.0, 120.0 | 90.0, 90.0, 120.0 |
| Resolution (Å) | 33.1–2.4 (2.48–2.40) | 56.1–2.8 (2.9–2.8) |
| $R_{meas}$ (%) | 8.3 (54.1) | 11.8 (78.2) |
| $I/\sigma$ (I) | 17.3 (2.4) | 5.0 (1.3) |
| Completeness (%) | 92.9 (86.1) | 99.2 (94.9) |
| Redundancy | 6.0 | 10.0 |
| Refinement | | |
| Resolution | 33.1–2.4 | |
| Number of unique reflections | 3097 | |
| $R_{work}/R_{free}$ | 20.9/25.3 | |
| Number of atoms | | |
| Protein | 1266 | |
| B-factors | | |
| Protein | 90.6 | |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.003 | |
| Bond angles (°) | 0.733 | |

Values in parenthesis are for the highest-resolution shell.

### Expression and solubility of P2 Cox variants

pEE720 derivatives containing the wild-type *cox* gene or various alanine substitutions were transformed into BL21(DE3) pLysE. The bacteria were grown in 35 ml of LB medium at 37°C until $OD_{600}$ 0.6 when IPTG was added to a final concentration of 0.5 mM. The culture was allowed to grow another 3 h at 37°C until harvested by centrifugation. To monitor the expression of the different Cox variants a small aliquot was collected before centrifugation and loaded onto 20% homogenous SDS–PAGE gels using the PhastSystem (GE Healthcare). The pellet was resuspended in Buffer C1 (0.3 M potassium phosphate buffer, pH 7.5, 3 mM EDTA, 0.5 M KCl, 0.1% Triton) at 6 ml per gram of cells and freeze thawed twice to allow leakage of lysozyme and partial lysis. To complete lysis, resuspended cells were sonicated on ice in two 15 s bursts at 12–14 μm with an MSE Soniprep 150. To decrease the viscosity of the lysate 10 units of DNaseI (Fermentas) was added and incubated for 5 min at 37°C. The extract was clarified by centrifugation in a Sorvall RC5C at 16 000*g* for 2 h at 4°C. Total protein concentration of the lysates was measured in a Nanodrop (Thermo Scientific). To monitor the solubility of the different Cox variants, 75 μg of total protein from the cleared lysate was loaded onto 17% ClearPAGE SDS gels (C.B.S.Scientific).

### Data analysis

An alignment of 20 cox-boxes (34) was used as input into WebLogo (35) to generate a sequence logo of the cox-boxes. All structural figures were prepared using PyMol.

## RESULTS

### Structure of P2 Cox

Crystallization experiments yielded crystals of P2 Cox in space group $P6_5$, diffracting to 2.4 Å (Table 1). The structure was solved with the SAD technique, using selenium-methionine incorporated Cox. The asymmetric unit contained one molecule of P2 Cox, and the final model comprises of residues 7–87. The first 6, the last 4 and the loop residues G50–R51 were (partially) disordered and not visible in the resulting electron density. The final $R_{work}/R_{free}$ values were 20.9/25.3%. The overall fold of P2 Cox shows a winged HTH structure consisting of a three-stranded antiparallel beta-sheet with a 12 residue loop packed against the HTH motif and the C-terminal domain containing two α-helices. The secondary structure

elements are arranged in a β1-α1-T-α2-β2-L-β3-α3-α4 order, with T and L being the turn and loop structures, respectively (Figure 1a).

The overall fold is similar to the previously determined $^{\lambda\Delta55}$Xis (20,12,13) (PDB code: 2IEF), with an r.m.s.d. of 1.0 Å over 40 Cα atoms between the two structures. The major difference is the loop between β2 and β3 in the wing, which in P2 Cox is much longer with 12 residues versus 3 in $^{\lambda}$Xis. In the $^{\lambda\Delta55}$Xis-DNA complex structure this loop binds to the minor groove of DNA. In contrast to the $^{\lambda\Delta55}$Xis protein, the full-length P2 Cox protein was crystallized, thereby revealing the structure of the C-terminal domain with helices α3 and α4 (Figure 1b).

Structural similarity is also found with the Xis protein of the conjugative transposon Tn916, $^{Tn916}$Xis (PDB code: 1Y6U), resembling the fold of Cox with an overall r.m.s.d. of ∼2 Å, and with a sequence similarity of 11% (22). Similar to $^{\lambda}$Xis, the loop of the wing in $^{Tn916}$Xis is much shorter compared to the one of Cox. The $^{Tn916}$Xis protein is thought to enhance excision in a similar manner to the $^{\lambda}$Xis protein, although clear mechanistic differences between the two have been suggested (23).

## P2 Cox oligomerization

It has previously been shown that oligomerization of P2 Cox is required for its biological activity (17). By analyzing the crystal packing we can show that extensive contacts are formed between the P2 Cox monomers in the crystal. Each monomer interacts with three other monomers to form a helix shaped packing in the crystal (Figure 1c). The helical filament has six monomers per turn, a diameter of 65 Å and a rise of 35 Å per turn, and is left-handed. The intermolecular interactions bury a large surface area; monomer $n$ has a buried surface area of 1040 Å$^2$ to molecule $n+1$, 1050 Å$^2$ to molecule $n-1$ and 245 Å$^2$ to molecule $n-2$. Since the total surface area of each monomer is ∼4300 Å$^2$, 55% is involved in intermolecular contacts. This is a very high percentage, normal values for biological oligomers are between 25% and 40% (21,36), strongly indicating that the packing is not an artifact from the crystallization but indeed biologically relevant.

The two C-terminal α-helices of P2 Cox contribute extensively to the helical packing. Helix 4 contains two tryptophans, W81 and W84, which both promote oligomerization. W81 forms a hydrophobic core together with several hydrophobic residues in α3 in the $n-1$ chain, as well as π-stacking to P38 of the same neighboring chain (Figure 1d and e). W84 has hydrophobic interactions with K34 and K36 on $n-1$, as well as with the D44 on $n-2$ (Figure 1d and e). Numerous hydrogen bonds and hydrophobic interactions along the interfaces also contribute to the packing of the proteins into the spiral shaped crystal arrangement. Interestingly, the DNA-binding motifs of P2 Cox, i.e. the wing binding to the minor groove and α-helix 2 binding to the major groove, do not significantly participate in the crystal packing, but are rather pointing outwards from the spiral (Figure 1c).

## Engineering of the putative interfaces of P2 Cox

Our P2 Cox structure indicates that α-helix 2 is the DNA-binding helix, similar to the homologs $^{\lambda}$Xis and $^{Tn916}$Xis. To further investigate the proposed DNA-recognition motif, amino acids thought to be involved in DNA binding were substituted for alanines, one at a time, by site-directed mutagenesis of plasmid pEE720 that contains the wild-type *cox* gene (Figure 2a). To test the biological activity of the mutations, the plasmids were transformed into strain C-6005, which is a lysogen containing the P2*cox3* mutated prophage (10). This P2 lysogen is defective in spontaneous phage production during growth (8). Upon transformation with plasmid expressing the wild-type P2 Cox protein, the spontaneous phage production increases and the number of plaques per colony forming bacteria (PFU/CFU) is about four orders of magnitude higher compared to the defective P2 lysogen. As shown in Figure 2b and Supplementary Table S2, most mutants give almost no release of phages at all, and thus are unable to complement the *cox3* defective prophage. The T25A, R30A, D33A and K75A mutants complement the defective prophage to some extent, although substantially less than wild-type P2 Cox. The loop of the wing structure of Cox is expected to contact the minor groove of DNA (Figure 1b). Therefore, R51 was also separately mutated into alanine. The substitution completely abolishes the biological activity in this system (Figure 2b), supporting the importance of the wing in DNA binding. Since the structure indicates that the C-terminal is involved in oligomerization and possible DNA interaction, amino acids W81 and W84, as well as R73, K75 and R78 were also substituted by alanine residues. Figure 2b shows that the mutants R73A R78A, W81A, W84A are all unable to complement the defective lysogen as opposed to K75A, which allows some spontaneous phage production. Finally, to confirm the importance of the last helix, a stop codon was inserted at position 70, and this eliminates P2 Cox protein activity. To ensure the different Cox variants do not greatly influence protein expression or folding, we overexpressed all variants and assayed their solubility. It was not feasible to do this in the complementation assay, as this assay relies on T7 promotor leakage, resulting in very low expression levels. All P2 Cox variants express, and all but two of the P2 Cox variants (D33A and P38A) express as soluble proteins as detected by SDS–PAGE on clarified cell extracts (Supplementary Figure S1). The D33A variant partially complements the defective prophage even though no soluble protein is detected. Likely there is still some soluble protein present, but at such levels that we do not see them by SDS–PAGE. In the case of P38A we cannot rule out that its lack of activity does not stem from its decreased solubility.

By comparing our P2 Cox structure with the $^{\lambda\Delta55}$Xis-DNA structure, we can couple mutations, both from the alanine scanning performed in this work as well as previously isolated mutants, to the presumed function of the amino acid substituted. An alanine exchange can in principle affect three things in P2 Cox: the DNA binding, the protein packing, the oligomerization or a combination thereof. Of the mutants shown in Figure 2b, nine are

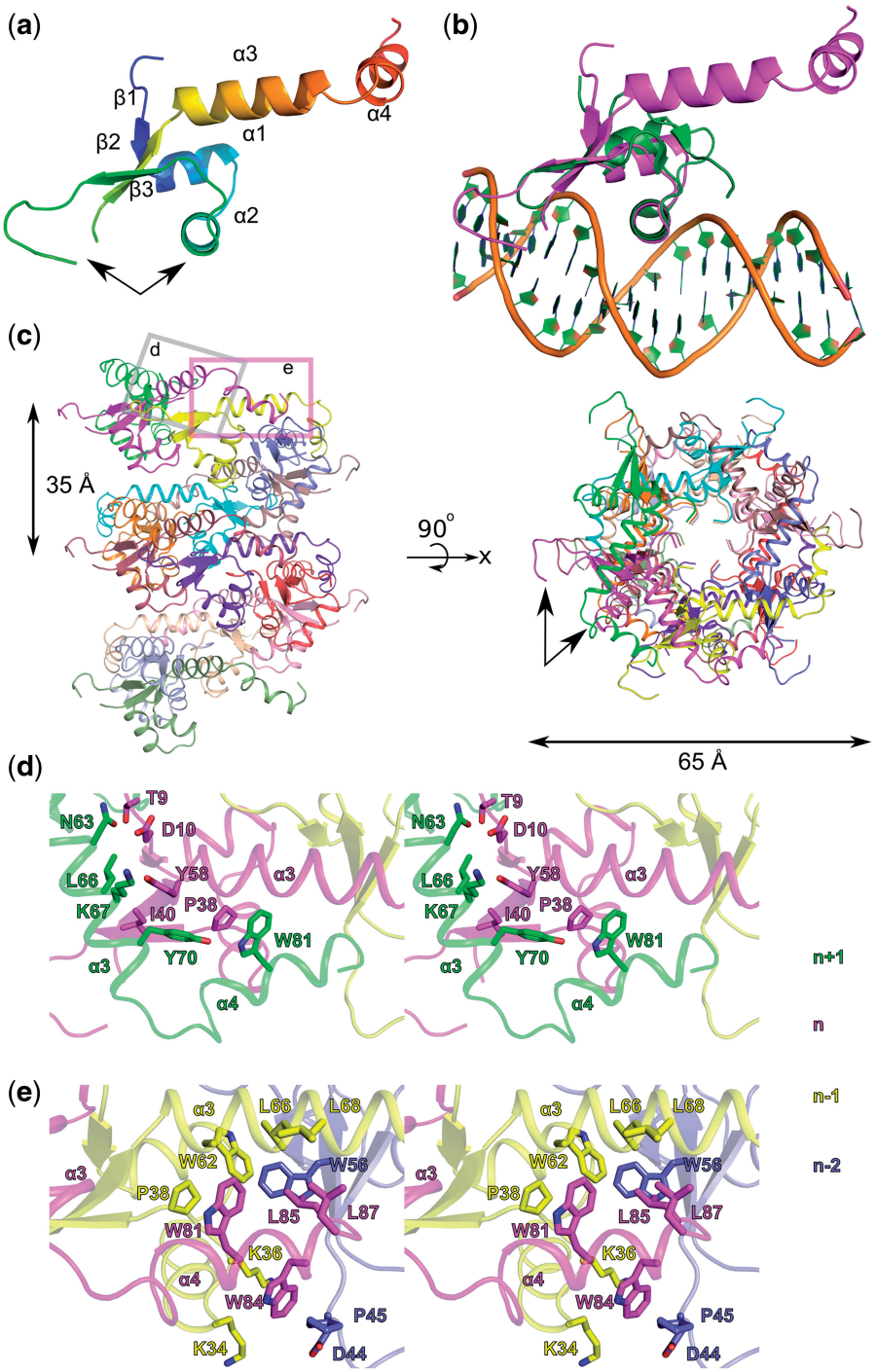


**Figure 1.** Structure of P2 Cox. **(a)** Cartoon representation, colored blue to red from the N- to the C-terminal. All secondary structure elements are labeled in the figure. The DNA-binding helix and wing are both marked with an arrow. **(b)** Superimposition of P2 Cox (magenta) and the DNA bound structure of $^{\lambda\Delta 55}$Xis (green, PDB code: 2OG0). The structural fold of the DNA interacting elements is conserved, with the exception of the much longer loop, interacting with the minor groove, in P2 Cox. **(c)** Visualization of the crystal packing of P2 Cox, from the side (left panel) and from the top (right panel). Cox packs into a left handed spiral, with a diameter of 65 Å and a pitch of 35 Å. Each protomer in the helical filament is colored individually. **(d** and **e)** Stereo figures of close-up views on the major oligomerization areas, with important residues shown as sticks. The approximate location of the close-up in respect to the helical filament is shown by the gray (d) and magenta (e) rectangles in (c) (~90° rotated). Four protomers in the helical filament are shown, $n+1$, $n$, $n-1$ and $n-2$ in green, magenta, yellow and blue, respectively.

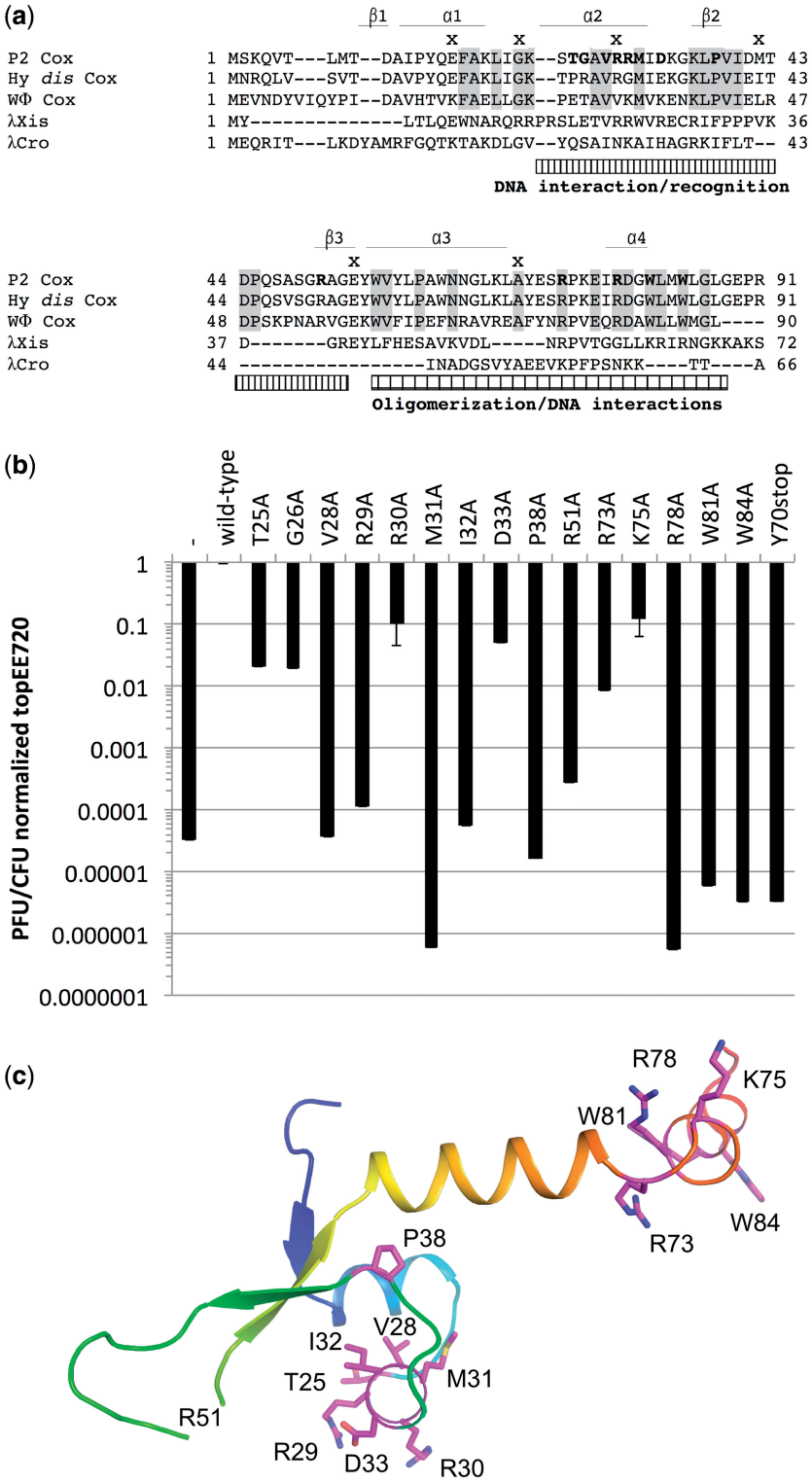


**Figure 2.** Functionally important residues in P2 Cox. **(a)** Sequence alignment of P2 Cox and its closest sequence and structural homologs. Single alanine substitutions of P2 Cox that have been examined within our work are in bold font. Residues marked with an x above are altered in some previously identified *cox* mutants. **(b)** Level of spontaneous phage production, measured via number of plaque forming units per colony forming unit. C-6005 is a *cox* deficient strain. All single alanine mutations were made on the previously designed pEE720 plasmid. All data have been normalized against the pEE720 wild type and are shown on a logarithmic scale. Error bars show the standard deviation of the normalized values over at least three independent measurements. For the raw data, please see Supplementary Table S2. **(c)** The positions of the residues that were exchanged into alanine in this study are shown as purple sticks in the cartoon representation of P2 Cox, colored blue to red from the N- to the C-terminal. R51 is located in a disordered loop, and its location is therefore only approximate.

suggested to directly affect the DNA binding (T25A, R29A, R30A, D33A, R51A, R73A, K75A, R78A and Δ70) to various degrees. Four of the alanine substitutions are suggested to disturb the protein packing: (i) V28A destroys the hydrophobic packing with α-helix 1, (ii) M31A will interfere with the hydrophobic packing of V28, L37 and K36. (iii) I32A could also lead to a reduced stability due to a decrease of hydrophobic interactions with Y14, L37, V39 and Y55. (iv) P38A could weaken the stability and rigidity provided to the backbone by the proline, possibly destabilizing the region preceding the wing motif and thereby also disturbing the DNA-binding capacity. Alternatively the effect could be due to decreased solubility of the P38A variant (Supplementary Figure S1). Finally, we identified five mutants of P2 Cox that likely affect the oligomerization, by interfering with either hydrophobic packing or stacking effects to other protomers within the oligomer. These are M31A (with $n-1$ chain), P38A (with $n+1$ chain), W81A (with $n-1$ chain), W84A (with $n-1$ chain) and Δ70 (with $n-1$ chain). W81A and W84A are located on the extended C-terminal, far from interactions with the winged HTH motif in the monomer. The exposed position of these large hydrophobic amino acids, usually found in the hydrophobic core of proteins, prompted us to investigate their importance by mutagenesis. The structure suggests that W81 and W84 are involved in P2 Cox oligomerization. The loss of function of the W81A and W84A variants point to the importance of these residues for protein function. Positions of the substituted residues in P2 Cox are shown in Figure 2c.

Several P2 Cox mutants have previously been isolated either as prophages unable to produce phages spontaneously, i.e. with impaired excision activity, or as P2 mutants unable to induce DNA replication of prophage P4, i.e. unable to act as a transcriptional activators (6,8,10,16,24). These P2 Cox mutations are shown in the sequence alignment in Figure 2a. The P2 Cox mutants identified as prophages unable to spontaneously form free phages during growth, i.e. *cox2* (M42T), *cox3* (G22E) and *cox4* (E54K), all have substitutions that probably disturb the protein packing as well as the oligomerization of the protein complex. Introduction of a charged glutamate in the place of a small glycine in a tight turn after α-helix 1, as in *cox3*, could impair this structural motif. The E54K exchange in c*ox4* produces a charge change on the amino acid side chain that likely extinguish charge interactions with the adjacent polypeptide chain in the oligomer. Among the *cox* mutants isolated as unable to induce prophage P4, *cox107* (R29H) suggestively affects the DNA-binding capacity. The A69T substitution in *cox129* introduces a polar residue in a hydrophobic patch including both chain $n+1$ and $n-1$, thereby potentially influencing the oligomerization behavior of P2 Cox. *Cox130* (E16K) does not substantially interfere with any interactions, neither within the P2 Cox packing nor in the oligomerization. This corresponds well with previous data which showed that this substitution does not influence P2 Cox activity in a substantial way (17).

## DISCUSSION

The Cox proteins constitute a unique family of directionality factors, coupling the control of the transcriptional switch with the integration/excision process (18,14,11). Phylogenetically P2 Cox belongs to a sub-group of this family, which has the predicted winged HTH domain at their N-terminal halves of the protein (37).

The crystal structure of P2 Cox presented in this article reveals that it shares the winged HTH with its homologs. Importantly, P2 Cox was crystallized as a full-length protein, in comparison to its homolog $^{\lambda}$Xis in which only the winged HTH domain was crystallized ($^{\lambda 55}$Xis). The structure reveals that P2 Cox has a defined structure of its C-terminal half, giving rise to both oligomerization and DNA-binding possibilities. Structural comparison of P2 Cox with the $^{\lambda 55}$Xis-DNA structure allows us to identify amino acids, in the winged HTH, that are positioned so that they can make either specific or non-specific contacts with the target DNA in the cox boxes. To confirm the biological importance of these residues, they were independently substituted by alanines, and checked for biological activity via a phage production assay. Most substitutions effectively inactivate P2 Cox. However, five mutants (T25A, G26A, R30A, D33A and K75A) are able to complement the *cox3* defective prophage, albeit at highly decreased levels, indicating that the exchange of these residues alone is not enough to completely disturb the protein function.

By analyzing previous biochemical data in the light of our new structural information, we can now propose which residues are important for DNA recognition. Two close homologs of P2 Cox have previously been studied, the Cox proteins from the phages P2 Hy *dis*, and WΦ (18,38,39). Hy *dis* Cox can complement a P2 *cox3* defective prophage, whereas WΦ Cox cannot. Sequence comparison, together with the present P2 Cox structure, indicates a couple of residues likely to be involved in the specific DNA base recognition such as T25 and D33. Both the T25A and the D33A mutants complement the defective prophage to some extent, although substantially less than wild-type Cox. Hy *dis* Cox, which can complement *cox3*, has a glutamate at position 33 while WΦ Cox has a lysine. This indicates that the negative charge in P2 Cox and Hy *dis* is important for DNA recognition. T25 is a threonine in both P2 and Hy *dis* Cox, whereas it is a proline in WΦ Cox. For other residues likely to be involved in DNA binding, no such dissimilarities can be observed between the three Cox proteins, thereby suggesting that T25 and D33 are important for DNA specificity.

Arginines are often found in interactions with DNA (40). We therefore also studied two arginines positioned in α-helix 2 (R29 and R30), as well as the well-conserved arginines found in the C-terminal close to α-helices 3 and 4 of P2 Cox (R51, R73 and R78). The alanine scanning showed that all but the R30A substitutions entirely abolish the biological function of the protein (Figure 2b). The three latter residues are all conserved between the Cox proteins discussed here, indicating that they are functionally important. R51 is located in the wing domain, which likely binds to the minor groove of the

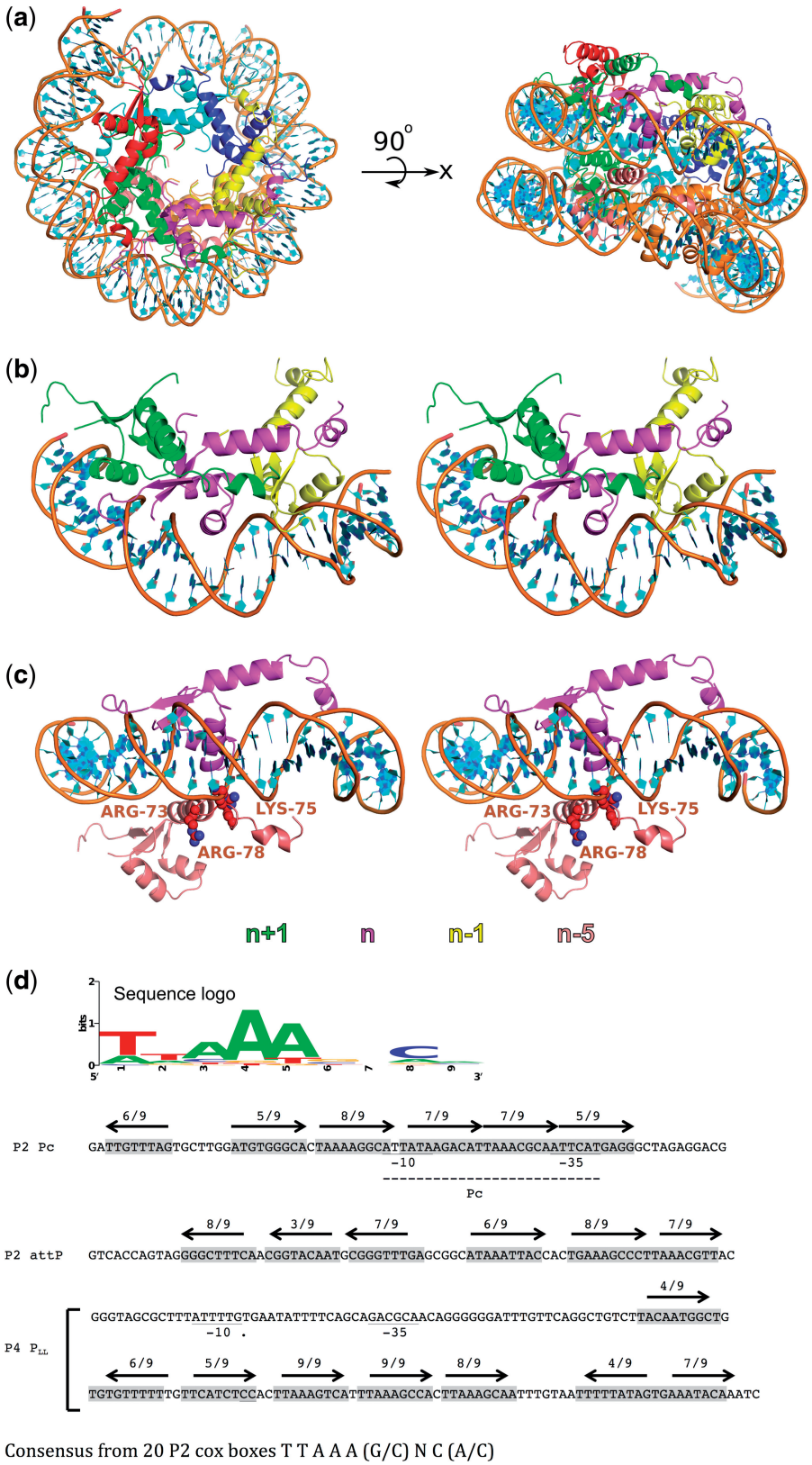


**Figure 3.** P2 Cox oligomerization and DNA binding. **(a)** Model of how P2 Cox can bind and wrap DNA around its oligomeric state, as seen from the top (left panel) and the side (right panel). The P2 Cox spiral has a similar diameter as DNA wrapping around nucleosome core particles, and the modeled DNA is taken from a DNA bound nucleosome structure (PDB code: 3AV1). **(b)** Stereo-view of a zoom in three of the protomers ($n+1$, $n$ and $n-1$) in the filament. Coloring scheme is the same as in Figure 1. To simplify the visualization of the DNA binding, only part of the modeled DNA from (a) is shown. The proposed DNA interacting elements of P2 Cox, the helix and wing, align well with the major and minor groove of the DNA, respectively, along the filament chain. **(c)** Stereo-view, rotated 45° around the $x$-axis when compared with (b). For clarity we are here only

(continued)

DNA (Figures 1b and 2c). R73 and R78 do not partake in any major interactions in the helical filament, rather they are pointing outwards in the helical filament. In a DNA-binding model, those residues of the $n-5$ protomer could interact with the backbone of the same stretch of DNA that is bound via the helix and wing of the n protomer (and $n-6$ with the DNA bound to $n-1$, etc) (Figure 3a–c). The importance of α-helices 3 and 4 is further supported by the fact that a deletion of the last 20 amino acids destroys the biological activity of P2 Cox.

An intriguing finding is the relatively large variation in spontaneous phage production in the complementation assay with some of the alanine substitutions (Supplementary Table S2). Since the experiments are standardized, and the controls do not show such large variations, it is most likely a result of an unknown factor affecting the transcriptional switch during growth. For example, the combination of the Cox3 protein expressed from the prophage and the alanine substituted Cox proteins expressed from the plasmids may form hetero-oligomeric structures with unknown functionality. Those could affect the DNA binding and consequently the downstream reactions in a stochastic way that give different outcomes in an overnight culture. However, the clarification of this is beyond the scope of this work.

In gel-filtration and cross-linking experiments Cox has been show to form tetramers and octamers in the absence of DNA (17). This indicates that Cox might initially associate to its DNA targets as a multimer *in vivo*. Gel shift analysis have been performed (17,41), and with increasing Cox concentration slower migrating DNA fragments are seen in the gels, indicating formation of larger protein complexes, but the number of Cox molecules per DNA fragment has never been determined. Thus, larger protein complexes are necessary for biological activity. Although we here show P2 Cox filament formation *in vitro* also in the absence of DNA, which happens at relatively high protein concentrations, we suggest that DNA binding may initiate and facilitate filament formation. By analysis of the spiral shaped crystal packing together with the biochemical data presented here we can show that the extensive contacts formed between the P2 Cox monomers in the crystal are likely to be important for the biological function. The contacts bury 55% of the surface area of the protein, strongly indicating that the helix-shaped crystal packing is not a crystallization artifact. Intriguingly, the diameter and pitch of the formed P2 Cox spiral is virtually the same to how DNA wraps around nucleosome core particles, which also have a diameter and pitch of $65 \pm 5$ and $30 \pm 5$ Å, respectively, as measured on different nucleosomes (PDB codes 3AV1, 1AOI, 1EQZ, 1F66, 1M1A and 2CV5). Furthermore, the P2 Cox helical filament is left-handed, as are nucleosomes (42). Although not a direct proof, it is very tempting to speculate that P2 Cox in a similar manner binds to its target DNA, by wrapping the DNA around its own spiral (Figure 3a). Previous biochemical data also supports this theory. It has been shown that P2 Cox upon binding induces DNA bending, but only if there are more than four cox-boxes (18). Furthermore, DNA footprinting experiments have shown that DNA bound to P2 Cox is sensitive to DNase at specific nucleotide distances, indicating DNA wrapping (8,9,18,23). Finally, the DNA-binding elements of P2 Cox are located favorably on the outside of the spiral, with ∼27 Å distance between the Cα atom of R30 and the Cα atom of R30 of any neighboring monomer in the filament, consistent with the distance between the major grooves of bent DNA (Figure 3b). This allows P2 Cox to interact with the DNA wrapped around the P2 Cox spiral (Figure 3a and b). The cox-boxes are distributed with 10–11 bases between the start of each site, consistent with the modeled P2 Cox–DNA complex, with one protomer binding to the major groove with each turn of the DNA helix (Figure 3b).

These results now allow us to add a molecular understanding to how P2 Cox functions in its three mechanistically very different functions: (i) repression of the P2 Pc promoter, (ii) activation of the P4 satellite phage and (iii) as an activator of excision. P2 Cox has been shown to protect a DNA region of at least 70 base pairs containing the different cox-boxes. At high P2 Cox concentrations the protected regions are extended around these regions (8,11,18). As can be seen in Figure 3d, the cox-boxes have different orientations in the different DNA target regions. We propose that the cox-boxes act as initiator sites for filament formation, allowing P2 Cox to bind and wrap the DNA around itself. We also propose that filament formation only occurs in the direction of the orientation of the cox-boxes. First, the P2 Pc promoter region is overlapped by the cox-boxes thereby allowing P2 Cox to bind and wrap the DNA effectively stopping transcription by blocking the −10 and −35 regions. Furthermore, at higher P2 Cox concentrations, footprint experiments have shown that the filament also starts to grow in the direction of the first cox-box toward the opposite facing Pe promoter, located ∼60 base pairs upstream of Pc, thus also repressing the Pe promoter. This notion is also supported by the fact that P2 Cox is autoregulating (5). Second, in the $P_{LL}$ promoter of the satellite phage P4 the first cox-box is located downstream and facing away from the −10 and −35 regions, hence leaving the promoter region accessible for transcription upon P2 Cox DNA binding and wrapping. Previous footprint

**Figure 3.** Continued
showing the *n* and the $n-5$ protomer. The $n-5$ protomer is located underneath the *n* protomer, almost a full turn of the filament. R73, K75 and R78 from the $n-5$ protomer are highlighted, as red spheres, showing the possibility for them to interact with the DNA backbone. **(d)** The sequence logo of a cox-box is shown, taken from 20 different cox-boxes (34). The cox-box consensus sequence is T T A A A (G/C) N C (A/C). The DNA sequences of the P2 Cox target DNA in phage P2 and satellite phage P4, as well as the sequence logo of the cox-boxes, are shown. The locations of the cox-boxes are indicated in gray, with the orientation of the cox-boxes and their identity to the consensus sequence indicated above the respective cox-box. Note that the cox-boxes with reverse orientation have their similarity to the sequence logo in its complementary reading frame. The location of the −10 and −35 regions in P2 Pc and in P4 $P_{LL}$ are indicated (6,16,43).

analysis supports this since the cleavage pattern is unchanged over the $P_{LL}$ promoter with increased protein concentration (16). This would indicate that filament formation only occurs in one direction at $P_{LL}$. Third, at the *attP* region the six identified cox-boxes are facing two different ways, thus indicating filament formation in both directions, also supported by footprint analysis (11). By growing toward the core and arm binding sites of P2 integrase, P2 Cox filament formation could have implications on intasome formation and consequently phage integration. The inhibitory effect of P2 Cox on integration is well known (8). Here, cooperative binding of P2 integrase and P2 Cox to *attP* probably reduce the amount of Cox needed *in vivo* to achieve this inhibitory effect. By a similar filament formation mechanism in *attL*, P2 Cox probably also determines the nature of the excisome in phage excision.

Previous data suggests that the concentration of P2 Cox determines to what extent the DNA is protected upon P2 Cox binding (11,16). Since expression of P2 Cox leads to repression of the P2 Pc promoter, excision activity, as well as activation of the P4 satellite prophage, it is clearly of great importance for the organism to tightly control the P2 Cox expression levels. However, further studies are needed to clarify the exact mechanism of P2 Cox regulation.

To summarize, we have here presented the crystal structure of full-length P2 Cox. We have through genetic and functional studies indicated the importance of key residues, suggested by our structural results, likely to be involved in oligomerization and DNA binding. Furthermore, the C-terminal region was identified as responsible for packing into helical filaments. The helical oligomerization revealed by the crystal structure of P2 Cox has given us the basis to interpret previous biochemical results at a structural level. It is the DNA binding, in particular the DNA wrapping around the P2 Cox helical oligomer, which in a direct way either blocks or unblocks the promoter regions around the different cox binding sites, thereby acting as either a repressor or activator.

## ACCESSION NUMBERS

The structure has been deposited in the Protein Data Bank, with accession code 4LHF.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the beamline scientists at ESRF, France, SLS, Switzerland, BESSY, Germany and MAX-lab, Sweden. Furthermore, we would like to congratulate Prof. Joe Bertani on his 90th birthday, 23 October, 2013, who in 1951 published the isolation of phages P1, P2 and P3 from a lysogenic *E. coli* strain.

## FUNDING

Swedish Research Council [2010-5200]; Wenner-Gren Foundations and the Swedish Foundation for Strategic research (to P.S.), by an EMBO Long Term Fellowship; Marie Curie Actions [EMBOCOFUND2010, GA-2010-267146] to R.P.-A.B., by Carl Tryggers foundation to E.H.-L. and P.S., by Magnus Bergvalls Foundation to P.S. and R.O. and by Marcus Borgströms Foundation to R.O. Funding for open access charge: Carl Tryggers foundation and the Swedish Research Council.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Bertani,G. and Six,E.W. (1988) The P2-like phages and their parasite, P4. In: Calender,R. (ed.), *The Bacteriophages*. Vol.2, Plenum Publishing Corp, New York.
2. Nilsson,A.S., Karlsson,J.L. and Haggård-Ljungquist,E. (2004) Site-specific recombination links the evolution of P2-like coliphages and pathogenic enterobacteria. *Mol. Biol. Evol.*, **21**, 1–13.
3. Nilsson,A.S. and Haggård-Ljungquist,E. (2006) The P2-like bacteriophages. In: Calender,R. (ed.), *The Bacteriophages*, 2nd edn. Oxford University Press, New York.
4. Massad,T., Skaar,K., Nilsson,H., Damberg,P., Henriksson-Peltola,P., Haggård-Ljungquist,E., Högbom,M. and Stenmark,P. (2010) Crystal structure of the P2 C-repressor: a binder of non-palindromic direct DNA repeats. *Nucleic Acids Res.*, **38**, 7778–7790.
5. Saha,S., Lundqvist,B. and Haggård-Ljungquist,E. (1987) Autoregulation of bacteriophage P2 repressor. *EMBO J.*, **6**, 809–814.
6. Haggård-Ljungquist,E., Kockum,K. and Bertani,L.E. (1987) DNA sequences of bacteriophage P2 early genes cox and B and their regulatory sites. *Mol. Gen. Genet.*, **208**, 52–56.
7. Ahlgren-Berg,A., Henriksson-Peltola,P., Sehlen,W. and Haggård-Ljungquist,E. (2007) A comparison of the DNA binding and bending capacities and the oligomeric states of the immunity repressors of heteroimmune coliphages P2 and WΦ. *Nucleic Acids Res.*, **35**, 3167–3180.
8. Saha,S., Haggård-Ljungquist,E. and Nordström,K. (1987) The Cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *EMBO J.*, **6**, 3191–3199.
9. Anderson,W.F., Ohlendorf,D.H., Takeda,Y. and Matthews,B.W. (1981) Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature*, **290**, 754–758.
10. Lindahl,G. and Sunshine,M. (1972) Excision-deficient mutants of bacteriophage P2. *Virology*, **49**, 180–187.
11. Yu,A. and Haggård-Ljungquist,E. (1993) The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. *J. Bacteriol.*, **175**, 7848–7855.
12. Bushman,W., Thompson,J.F., Vargas,L. and Landy,A. (1985) Control of directionality in lambda site specific recombination. *Science*, **230**, 906–911.
13. Moitoso de Vargas,L. and Landy,A. (1991) A switch in the formation of alternative DNA loops modulates lambda site-specific recombination. *Proc. Natl Acad. Sci. U.S.A.*, **88**, 588–592.
14. Lewis,J.A. and Hatfull,G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.*, **29**, 2205–2216.
15. Six,E.W. and Lindqvist,B.H. (1978) Mutual derepression in the P2-P4 bacteriophage system. *Virology*, **87**, 217–230.
16. Saha,S., Haggård-Ljungquist,E. and Nordström,K. (1989) Activation of prophage P4 by the P2 Cox protein and the sites of action of the Cox protein on the two phage genomes. *Proc. Natl Acad. Sci. U.S.A.*, **86**, 3973–3977.

17. Eriksson,J. and Haggård-Ljungquist,E. (2000) The multifunctional bacteriophage P2 Cox protein requires oligomerization for biological activity. *J. Bacteriol.*, **182**, 6714–6723.
18. Ahlgren-Berg,A., Cardoso-Palacios,C., Eriksson,J.M., Mandali,S., Sehlen,W., Sylwan,L. and Haggård-Ljungquist,E. (2009) A comparative analysis of the bifunctional Cox proteins of two heteroimmune P2-like phages with different host integration sites. *Virology*, **385**, 303–312.
19. Sam,M.D., Papagiannis,C.V., Connolly,K.M., Corselli,L., Iwahara,J., Lee,J., Phillips,M., Wojciak,J.M., Johnson,R.C. and Clubb,R.T. (2002) Regulation of directionality in bacteriophage lambda site-specific recombination: structure of the Xis protein. *J. Mol. Biol.*, **324**, 791–805.
20. Abbani,M.A., Papagiannis,C.V., Sam,M.D., Cascio,D., Johnson,R.C. and Clubb,R.T. (2007) Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc. Natl Acad. Sci. U.S.A.*, **104**, 2109–2114.
21. Albright,R.A. and Matthews,B.W. (1998) Crystal structure of lambda-Cro bound to a consensus operator at 3.0 A resolution. *J. Mol. Biol.*, **280**, 137–151.
22. Abbani,M., Iwahara,M. and Clubb,R.T. (2005) The structure of the excisionase (Xis) protein from conjugative transposon Tn916 provides insights into the regulation of heterobivalent tyrosine recombinases. *J. Mol. Biol.*, **347**, 11–25.
23. Connolly,K.M., Iwahara,M. and Clubb,R.T. (2002) Xis protein binding to the left arm stimulates excision of conjugative transposon Tn916. *J. Bacteriol.*, **184**, 2088–2099.
24. Eriksson,S.K., Liu,T. and Haggård-Ljungquist,E. (2000) Interacting interfaces of the P4 antirepressor E and the P2 immunity repressor C. *Mol. Microbiol.*, **36**, 1148–1155.
25. Studier,F.W. and Moffatt,B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**, 113–130.
26. Collaborative Computational Project 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
27. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
28. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
29. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
30. Murshudov,G.N., Vagin,A.A. and Dodson,E. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
31. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
32. Sasaki,I. and Bertani,G. (1965) Growth abnormalities in Hfr derivatives of *Escherichia coli* strain C. *J. Gen. Microbiol.*, **40**, 365–376.
33. Sunshine,M.G., Thorn,M., Gibbs,W., Calendar,R. and Kelly,B. (1971) P2 phage amber mutants: characterization by use of a polarity suppressor. *Virology*, **46**, 691–702.
34. Eriksson,J. (2005) In: Haggård-Ljungquist,E. (ed.), *Structure-function Studies of Bacteriophage P2 Integrase and Cox Protein*. Department of Genetics, Microbiology and Toxicology, Stockholm University, Stockholm.
35. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
36. Wang,Y., Huang,Y., Wang,J., Cheng,C., Huang,W., Lu,P., Xu,Y.-N., Wang,P., Yan,N. and Shi,Y. (2009) Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature*, **462**, 467–472.
37. Nilsson,H., Cardoso-Palacios,C., Haggård-Ljungquist,E. and Nilsson,A.S. (2011) Phylogenetic structure and evolution of regulatory genes and integrases of P2-like phages. *Bacteriophage*, **1**, 207–218.
38. Liu,T. and Haggård-Ljungquist,E. (1999) The transcriptional switch of bacteriophage WΦ, a P2-related but heteroimmune coliphage. *J. Virol.*, **73**, 9816–9826.
39. Renberg-Eriksson,S.K., Ahlgren-Berg,A., DeGrooth,J. and Haggård-Ljungquist,E. (2001) Characterization of the developmental switch region of bacteriophage P2 Hy *dis*. *Virology*, **290**, 199–210.
40. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
41. Frumerie,C., Sylwan,L., Ahlgren-Berg,A. and Haggård-Ljungquist,E. (2005) Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. *Virology*, **332**, 284–294.
42. Luger,K., Dechassa,M.L. and Tremethick,D.J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.*, **13**, 436–447.
43. Dehó,G., Zangrossi,S., Ghisotti,D. and Sironi,G. (1988) Alternative promoters in the development of bacteriophage plasmid P4. *J. Virol.*, **62**, 1697–1704.