# 4. Results

I have determined the X-ray structure of the DNA-binding domain (KorB-O) of the transcriptional repressor and ParB homologue KorB in complex with a 17-bp oligonucleotide containing the operator ($O_B$) sequence. The atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession number 1R71.

A detailed description of various steps, which led to the successful determination of KorB-O—$O_B$ structure is given below. Further, the structure is explained in detail with respect to the protein fold and $O_B$ binding.

## 4.1. Choice of oligonucleotides

The KorB protein recognizes and binds specifically to the palindromic operator, $O_B$ (consensus sequence 5′ **TTTAGC$^{G}$/$_{C}$GCTAAA** 3′). For a successful protein-DNA cocrystallization experiment the choice of synthetic oligonucleotide is critical in terms of length and biological specificity. In this work, oligonucleotides of 13 and 17 bp length (Table 3) were tested for their KorB-O binding in a simple band shift assay (Fig. 8). As seen from the native gel in Fig. 8 a slight shift was seen for KorB-O in presence of a 17 bp $O_B$ fragment, whereas no visible shift could be seen for the 13 bp $O_B$. Based on this result, 17 bp $O_B$ was selected for further cocrystallization experiments.

**Table 3: Oligonucleotides derived from the consensus sequence and tested for their KorB-O binding**

| Oligonucleotide | Sequence |
|---|---|
| **13 bp $O_B$** | **5′ TTTAGC G GCTAAA 3′**<br>**3′ AAATCG C CGATTT 5′** |
| **17 bp $O_B$** | **5′** CT **TTTAGC C GCTAAA** AT **3′**<br>**3′** GA **AAATCG G CGATTT** TA **5′** |

shaded region is the 13 base pair consensus sequence

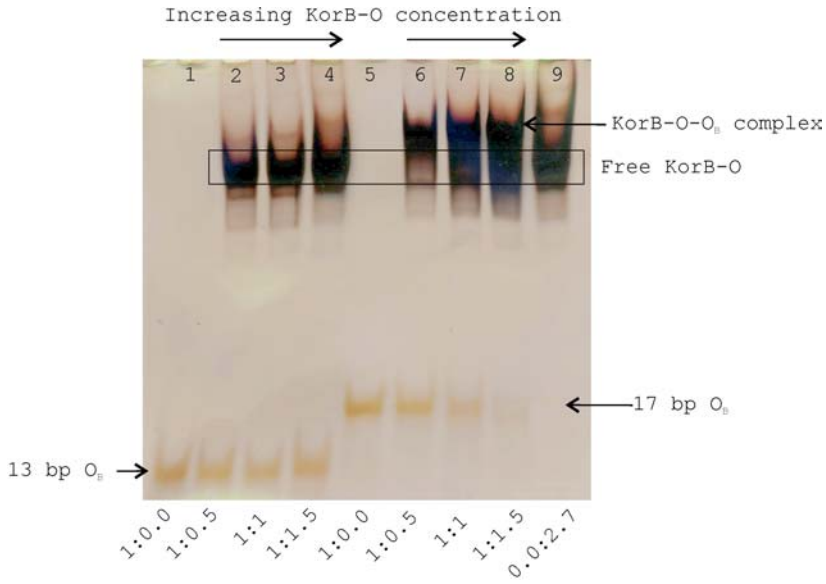



**Figure 8. Band shift assay for determining the binding of oligonucleotides with KorB-O.** All samples were incubated at 4 °C for 30 min before running on a 15 % native gel, which was subsequently stained with silver nitrate. Binding of KorB-O to 17 bp $O_B$ causes a shift that is seen in lanes 6 to 8. One can see a relative decrease in amounts of free DNA when the amount of protein is increased. All ratios are for DNA:Protein. Lanes 1 and 5 have free 13 bp $O_B$ (0.18 nanomoles/µl) and 17 bp $O_B$ (0.16 nanomoles/µl). Lane 9 contains free KorB-O (0.43 nanomoles/µl).

## 4.2. Phase problem with crystals of KorB-O and 17 bp $O_B$

A complex of KorB-O with the 17 bp $O_B$ was successfully crystallized. Crystals were obtained under a wide variety of conditions but the best ones diffracted to less than 3 Å at the synchrotron source. The next step was the determination of the phase angles for each of the diffracted intensities. For this, heavy atoms were screened, since molecular replacement could not be used. Unfortunately, neither the heavy atom derivatization nor quick soaking with NaBr (Dauter and Dauter, 2001) was able to give any interpretable electron density map. So another approach (brominated oligonucleotides) had to be taken.

## 4.3. Brominated oligonucleotides

To create brominated oligonucleotides, the 17 bp $O_B$ with previously confirmed binding to KorB-O was used. This oligonucleotide was brominated at three positions outside of the consensus sequence. This was done by chemical synthesis where three thymines were exchanged with 5-bromodeoxyuridine ($^{Br}U$) (Table 3 and Fig. 9). This brominated oligonucleotide is referred to further as $O_B$. A stable complex of KorB-O and $O_B$ (KorB-O—$O_B$) was obtained, purified and subjected to crystallization. $O_B$ was successfully used to solve the crystal structure presented here.

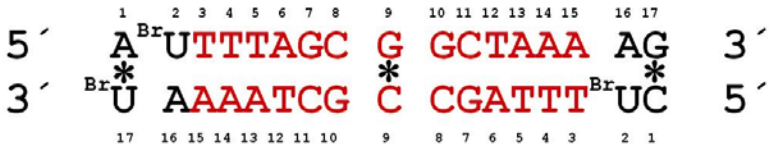



**Figure 9. Brominated oligonucleotide $O_B$ used for solving the crystal structure of KorB-O—$O_B$ complex**. The consensus sequence is given in red. Base pairs marked with asterisk (∗) break two-fold symmetry of the DNA duplex. The 5-bromo-deoxyuracil base is represented by $^{Br}U$.

## 4.4. Purification of the KorB-O—$O_B$ complex

The low concentration of KorB-O was critical when mixing with the oligonucleotides, as higher concentrations caused precipitates to form. The components were mixed by adding $O_B$ (16.6 nanomoles) to the KorB-O solution (27.7 nanomoles) in a 2:1.2 (KorB-O:$O_B$) molar ratio, and subjected to gel filtration (Superdex 75) to separate the protein-DNA complex from free DNA and/or protein (Fig. 10A). The KorB-O—$O_B$ complex migrates with a relative molecular mass of ~50 kDa, which corresponds to two KorB-O molecules bound at the $O_B$. The complex was concentrated to 18.6 mg/ml corresponding to an $A_{260}$ of 78.

A comparison of elution profiles from different runs of KorB-O—$O_B$, KorB-O and $O_B$ is also shown in Fig. 10B. A clear shift is seen in the complex peak (black) when compared with the elution pattern of protein or DNA alone. So the complex could be purified and separated from free KorB-O or free $O_B$.

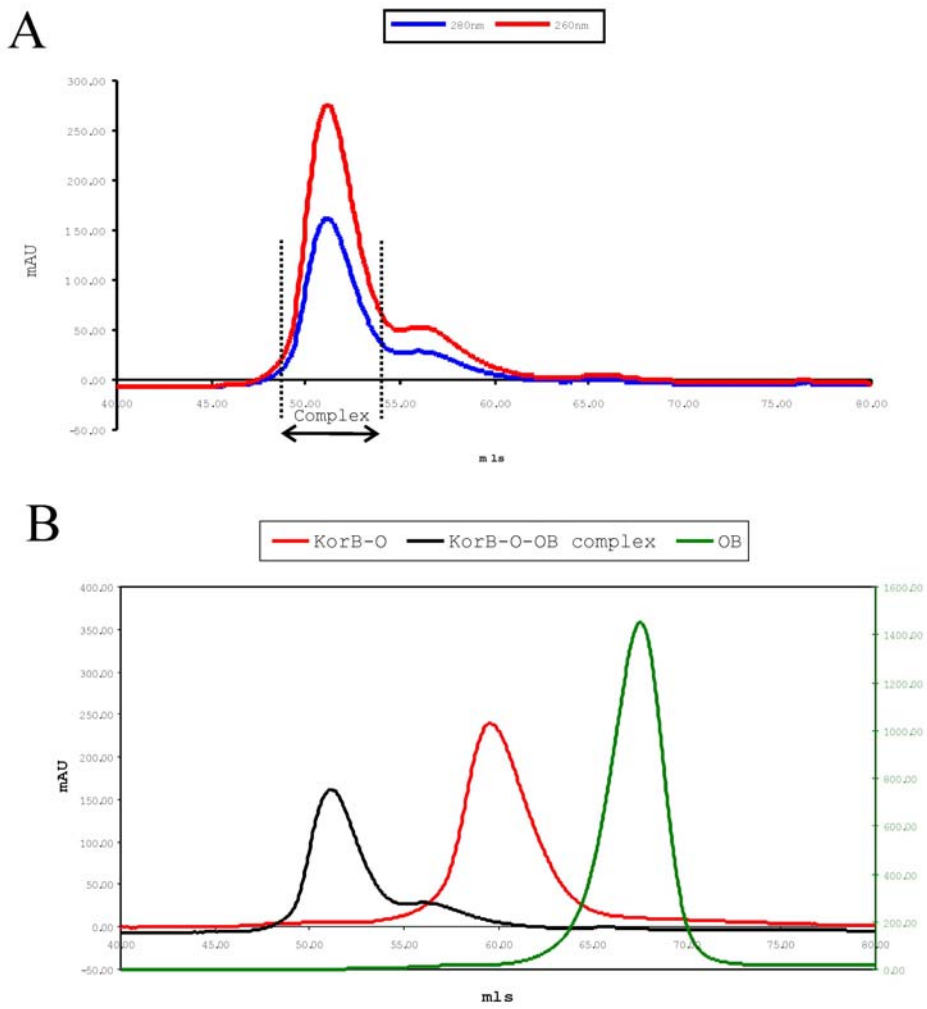



**Figure 10. Migration behaviour of the KorB-O—$O_B$ complex, KorB-O and $O_B$.** (A) Purification of the KorB-O—$O_B$ complex from free DNA or protein in solution over Superdex 75 gel filtration column. The complex migrates as a single peak with an apparent molecular mass of ~50 kDa molecular weight, when compared with a calibration curve for the column with known molecular weight standards. (B) The KorB-O—$O_B$ complex (black) elutes earlier when compared to the elution of free KorB-O (red) or free $O_B$ (green). The milli absorption units (mAU) are at $OD_{280}$ nm and a separate scale for the duplex DNA (green) is given on the right side of the figure 10B. All runs were carried out separately on a Superdex 75 (16/60) gel filtration column connected to an Äkta Explorer FPLC system. Different runs were made under similar buffer conditions (20 mM Tris.Cl pH 7.6 and 50 mM NaCl) and flow rate.

This final polishing step over the gel filtration column (Fig. 10) was important for crystallization success and the quality of the crystals.

## 4.5. Chemical crosslinking

KorB-O behaved as a monomer on a gel filtration column (Fig. 10). To see the effect of $O_B$ on KorB-O a simple chemical crosslinking experiment based on glutaraldehyde was performed as shown in Fig. 11. No higher order multimer was seen upon glutaraldehyde treatment of KorB-O alone. But quite interesting is the fact that in presence of $O_B$, glutaraldehyde is able to crosslink two KorB-O monomers and the effect is easily seen on a SDS PAGE gel in Fig. 11B.

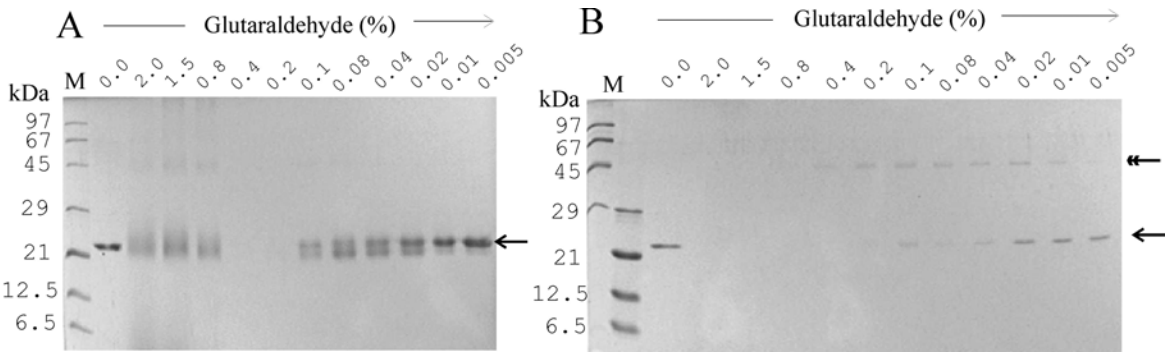



**Figure 11. Chemical crosslinking to see the oligomeric nature of KorB-O.** (A) KorB-O in presence of the crosslinker glutaraldehyde and (B) KorB-O—$O_B$ complex in presence of glutaraldehyde on a Coomassie stained 18 % SDS PAGE gel. Arrow near the 21 kDa marker marks the monomeric form of KorB-O, which has a molecular weight of 19.9 kDa. Higher concentration of glutaraldehyde (0.4%) causes the precipitation of KorB-O; these higher order aggregates are thereby not visible as they are not able to enter the gel. KorB-O runs a bit unusual on a SDS PAGE gel probably because of its low pI of 4.73. The double arrow near 45 kDa marker marks the dimeric form of KorB-O.

## 4.6. Crystallization of the KorB-O—$O_B$ complex

Trigonal crystals were obtained in spacegroup $P3_221$ ($a$ = 110.44 Å, $c$ = 160.53 Å), using the hanging-drop vapour-diffusion method at 20°C when the reservoir contained 800 µl of 25% MPD and 0.4 M ammonium dihydrogen phosphate (Fig. 12). Crystals grew in three days from a drop containing 1 µl of concentrated complex and 1 µl of the reservoir.

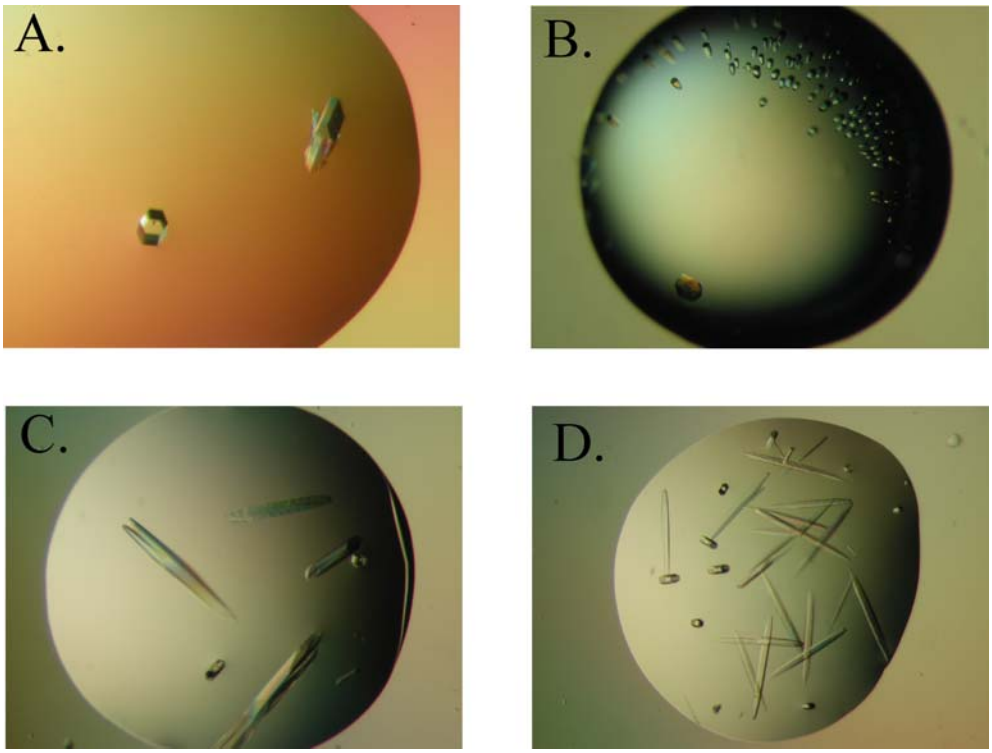



**Figure 12. Crystals of KorB-O—$O_B$ complex**. A. Crystals of the complex with $O_{B17}$ (non brominated DNA) were obtained using Screen I (Hampton Research, Aliso Viejo, USA), they were screened for heavy atom derivatives. B, C and D contain KorB-O in complex with 5-Br-dU labelled DNA. B. Initial crystals of the complex which were obtained under similar conditions as in A (without a cryoprotectant). Drop A, C and D crystals are with the cryoprotectant MPD. Drop C shows some crystals that are very long, but they diffract only to about 5 Å and show high mosaicity. Also the drop D contains the crystal (approximate dimensions 500×100×100 µm$^3$) that diffracted to 3 Å at the home source and was later used for collecting the MAD data at the synchrotron.

## 4.7. Data collection and processing

A crystal which diffracted to 3 Å at the home source was frozen in liquid nitrogen and a three-wavelength MAD experiment was carried out at 100 K using synchrotron radiation at the ID14.2 beamline at BESSY, Berlin, using a MAR345IP detector. As the anomalous scattering factors are derived from the atomic absorption coefficients, a fluorescence scan near the bromine edge (Fig. 13A) was carried out to find out the precise wavelengths for data collection. The choice of wavelengths for the peak ($\lambda_1$), inflection point ($\lambda_2$) and remote data ($\lambda_3$) was made so as to obtain the largest value for the anomalous difference, $\Delta f''$ ($\lambda_1$=0.92022 Å) and the minimum value for $f'$ ($\lambda_2$=0.92039 Å) (Fig. 13B). The far remote wavelength ($\lambda_3$=0.89844 Å) maximized the dispersive difference, $\Delta f'$($\lambda_3$-$\lambda_2$). The energies, $f'$ and $f''$ values are summarized in Table 4.

**Table 4. Energies for the three wavelength MAD experiment**

| | Energy (eV) (collected) | $f'$ | $f''$ |
|---|---|---|---|
| **Peak** | 13473.5 | -7.8 | 5.0 |
| **Inflection point** | 13471.0 | -10.0 | 2.6 |
| **High energy remote** | 13800.0 | | |

$f'$ and $f''$ values were derived using the program CHOOCH

(Evans and Pettifer, 2001)

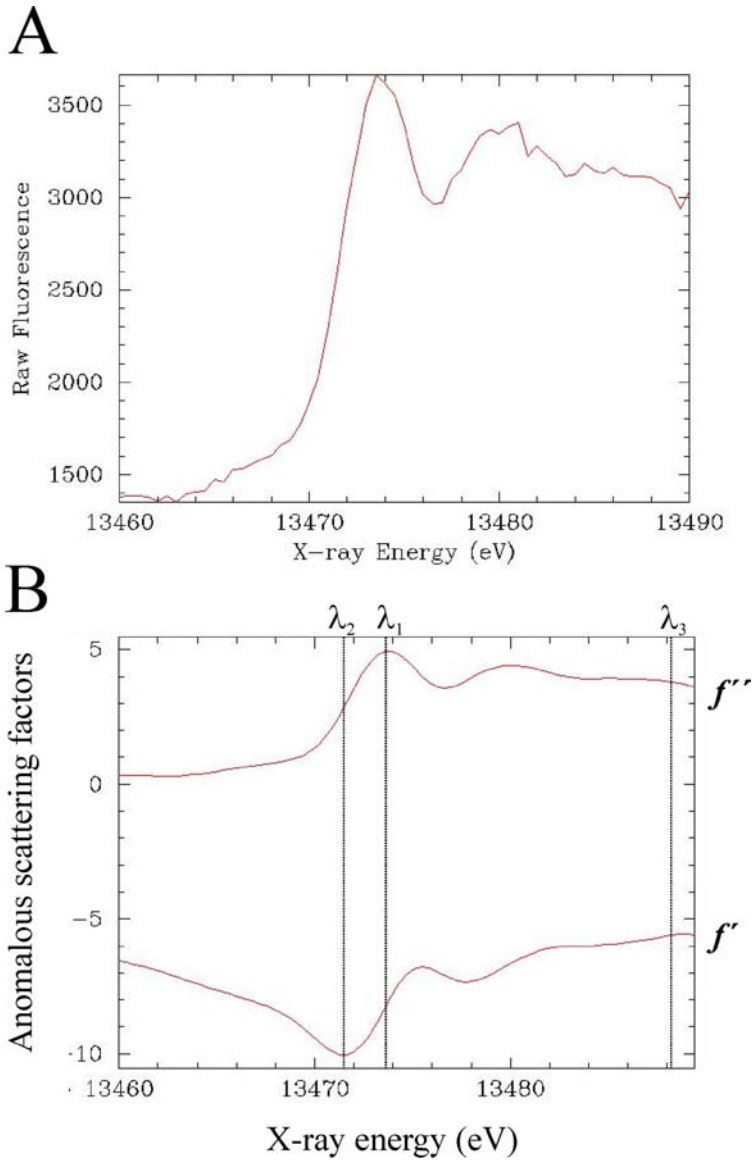



**Figure 13. An X-ray absorption spectrum obtained from 5-BrdU substituted oligonucleotide**. (**A**) The raw fluorescence spectrum was measured at ID 14.2, BESSY, Berlin, using a detector from Amptek, MA, USA. (**B**) Experimental determination of *f′* and *f′′* values. Anomalous scattering curves were derived using the program *CHOOCH* using the raw fluorescence data shown in the above figure.

For each wavelength a complete set of X-ray diffraction data were collected, each exposure involving 1° rotation of the crystal. About 200° of data were collected for each wavelength ($\lambda_1$ 205°, $\lambda_2$ 208° and $\lambda_3$ 210°). The XDS package (Kabsch, 1993) was used to integrate reflection intensities. The quality of the datasets is summarized in Table 5. A diffraction pattern from the X-ray data collection is shown in Fig. 14.

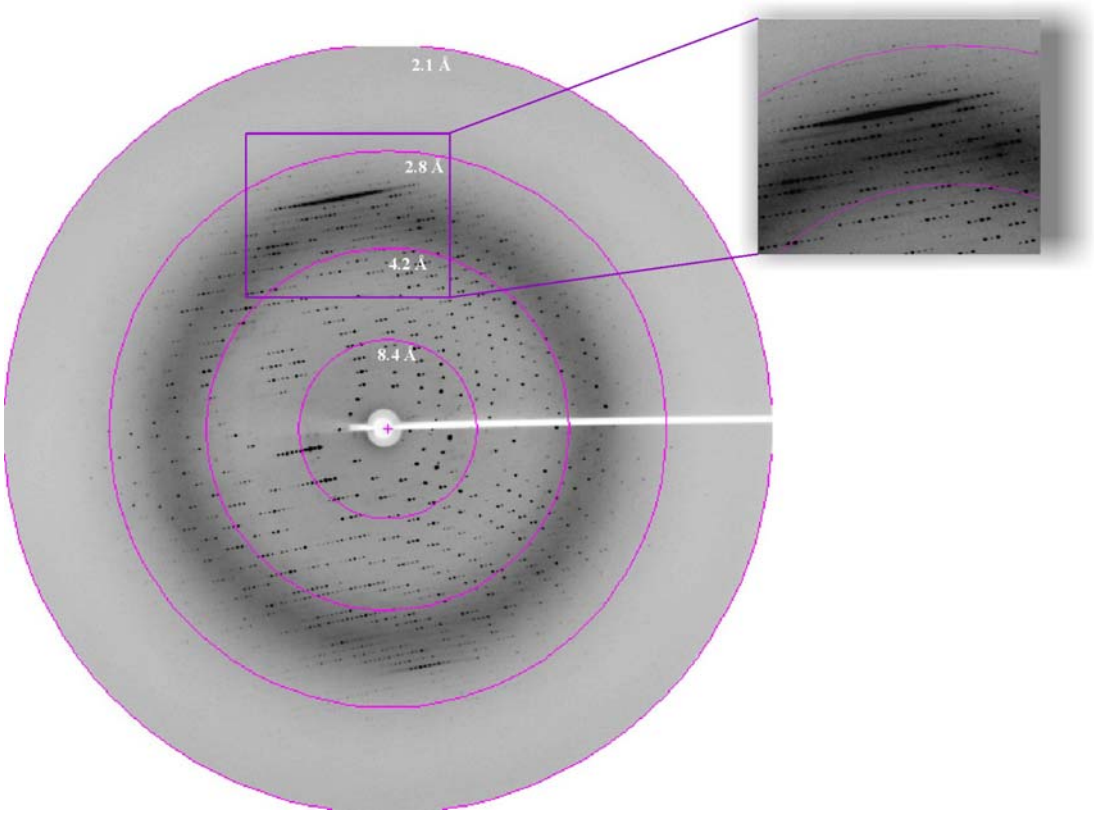



**Figure 14. Diffraction image from KorB-O—$O_B$ crystal (resolution 2.2 Å), measured at ID 14.2, BESSY, Berlin**. Although not apparent from the image, diffraction spots of low intensity are present in the outermost ring. Crystal to detector distance is 250 mm.

## Table 5. X-ray data and refinement statistics

| **Crystallographic data** | | | |
|---|---|---|---|
| Data set | Peak | Inflection point | High-energy remote |
| Wavelength (Å) | 0.92022 | 0.92039 | 0.89844 |
| Resolution range (Å) | 30-2.2 (2.4-2.2) | 30-2.5 (2.6-2.5) | 30-2.5 (2.6-2.5) |
| Total reflections[a] | 718,078 | 495,719 | 511,888 |
| Unique reflections[a] | 108,306 | 73,321 | 75,097 |
| Completeness (%) | 97.4 (92.9) | 96.8 (91.0) | 99.1 (97.7) |
| $< I / \sigma(I) >$ | 14.9 (3.9) | 12.3 (2.6) | 20.7 (7.2) |
| $R_{sym}$ (%) | 7.7 (49.2) | 12.3 (77.9) | 6.1 (26.0) |
| $R_{meas}$ (%)[b] | 8.3 (53.3) | 13.3 (84.3) | 6.6 (28.1) |

| **Refinement** | | *Ramachandran plot regions (%)*[c] | |
|---|---|---|---|
| Resolution range (Å) | 30-2.2 | Most favoured | 95.1 |
| $R_{work}$ (%) | 19.5 | Additionally allowed | 4.9 |
| $R_{free}$ (%) | 25.0 | Generously allowed | 0.0 |
| Reflections (work / free) | 54,758 / 2,911 | Disallowed | 0.0 |
| rmsΔ bonds (Å) | 0.015 | | |
| rmsΔ angles (º) | 2.012 | | |
| No. of protein and DNA atoms | 6422 | | |
| No. of water oxygens | 347 | | |

Values in highest resolution shell are given in parentheses.

[a] Friedel pairs not merged.

[b] Multiplicity corrected $R_{sym}$ as defined by Diederichs and Karplus, 1997.

[c] Values calculated using PROCHECK (Laskowski *et al*., 1993).

The KorB-O—$O_B$ complex crystallized in spacegroup $P3_221$ with cell parameters $a$ = 110.44 Å, $c$ = 160.53 Å, with two copies of the 49.8-kDa complex in the asymmetric unit and with a solvent content of ~60 %.

## 4.8. Phasing, model building and structure refinement

Structure determination of the KorB-O—DNA complex was based on covalent modification of the DNA with bromine at three positions outside the consensus operator sequence where 5-bromouracil was substituted in place of thymines (Fig. 9). Experimental crystallographic phases were obtained using the MAD data recorded around the bromine edge. MAD data (30-2.5 Å) were used in SOLVE (Terwilliger and Berendzen, 1999) to obtain 6 of the bromine sites, yielding an overall figure of merit of 0.43. The program RESOLVE (Terwilliger, 2000) was used for solvent flattening which led to an overall figure of merit of 0.60 and the automatic modeling of 213 protein residues out of 712 with side chains.

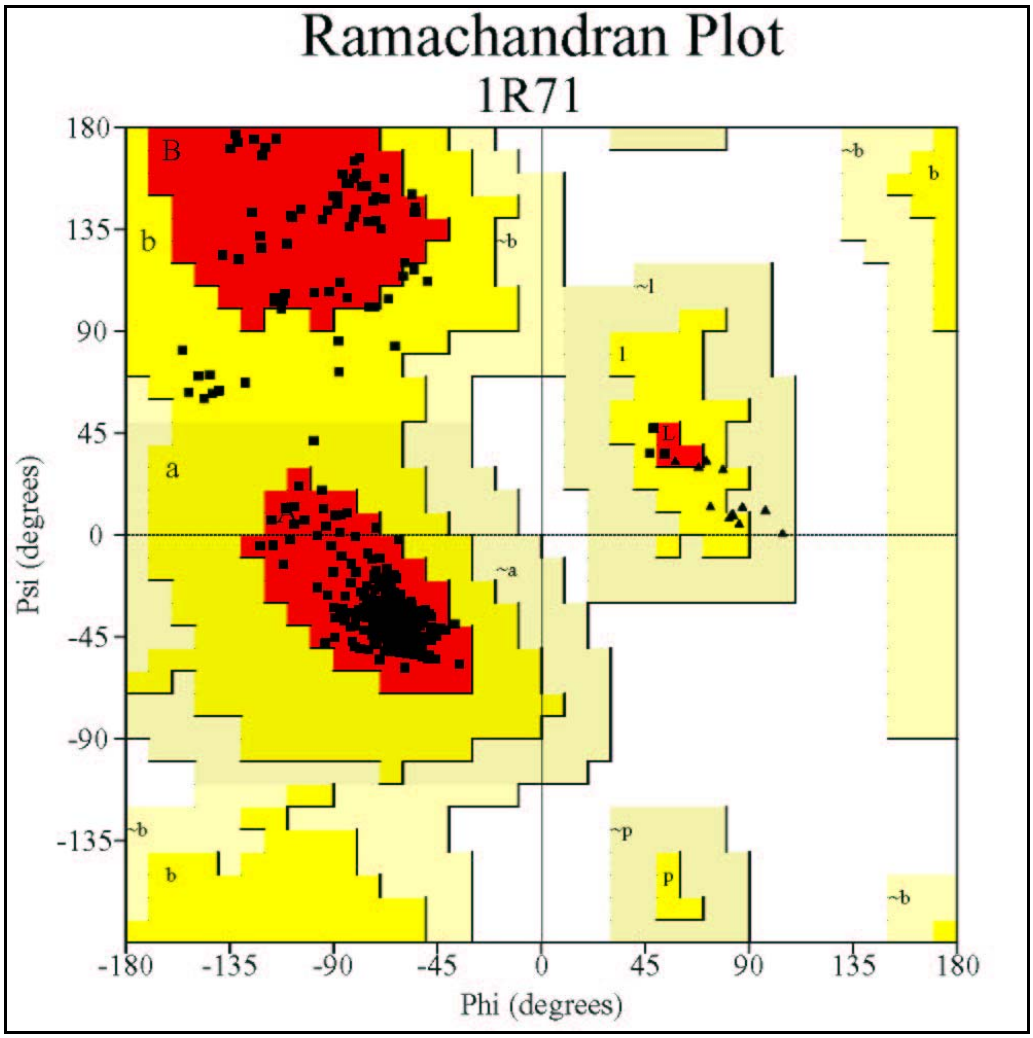
An electron-density map was obtained at 2.5 Å resolution where the DNA was easily recognized as a straight B-form double helix. A DNA model was manually fitted as a rigid body into the electron density taking the bromine positions as pivot points, using the program O (Jones *et al*., 1991). Prior to refinement, 5% of the data was kept aside for cross-validation, and $R_{free}$ was used to monitor model refinement strategy and progress (Brunger, 1992). Rigid-body refinement was carried out with the REFMAC5 (CCP4, 1994) program resulting in a decrease in $R$ value from 46 to 37%. Further refinement was carried out using the restrained refinement procedure of maximum-likelihood residual of REFMAC5. Residues were built in the $2mF_o$-$DF_c$ electron-density map, and after successive cycles of refinement the $R_{work}$ dropped to 26.3% and the $R_{free}$ to 30.4%.

At this stage I started using the peak data set where the resolution extends to 2.2 Å. A full unique set of reflections was created using the CCP4 suite, $R_{free}$ labels were imported from the experimental phases and extended to 2.2 Å, and now the phases from the model were used. Water molecules having chemically reasonable distances to potential hydrogen-bond donors/acceptors, were added to the model using both automated methods (ARP/wARP) (Perrakis *et al*., 2001) and manual inspection of difference maps. Towards the end of refinement, strong positive density was observed in a $F_o$-$F_c$ map at the $N^7$ position of G17

(Fig. 20), and the central and terminal base pairs looked as though there was some conformational averaging. This was accounted for by superimposing the double-stranded DNA in two orientations according to its pseudo-dyad axis and adjusting the occupancy to 0.5 for all atoms. After a few cycles of restrained refinement the positive density disappeared and the $R_{free}$ dropped from 26.3% to 25.1%. The Fourier maps are of good quality with all DNA bases and backbone defined and represent the average electron density of the two DNA orientations (illustrated section 4.11). Final refinement steps resulted in $R_{free}$ and $R_{work}$ of 25.0 and 19.5%, respectively.

## 4.9. Validation

Model quality was checked with WHATCHECK (Hooft *et al*., 1996) and PROCHECK (Laskowski *et al*., 1993). A PROCHECK and PROMOTIF (Hutchinson and Thornton, 1994) summary for the structural motifs of KorB-O is given in Appendix VI and VII, respectively. The quality of macromolecular structure factor data and their agreement with the atomic model was checked with the program SFCHECK (Vaguine *et al*., 1999), which yielded a correlation factor of 0.929 (Appendix VIII). The KorB-O model has good stereochemistry with no outliers and 95.1% of the non-glycine residues falling in the most favoured regions of the Ramachandran diagram (Ramakrishnan and Ramachandran, 1965) (Fig. 15 and Table 5).




**Plot statistics**

| | | |
|---|---|---|
| Residues in most favoured regions [A,B,L] | 389 | 95.1% |
| Residues in additional allowed regions [a,b,l,p] | 20 | 4.9% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 0 | 0.0% |
| Residues in disallowed regions | 0 | 0.0% |
| | ---- | ------ |
| Number of non-glycine and non-proline residues | 409 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 20 | |
| Number of glycine residues (shown as triangles) | 24 | |
| Number of proline residues | 16 | |
| | ---- | |
| Total number of residues | 469 | |

**Figure 15. Ramachandran plot.** Non-glycine residues are depicted as filled squares; filled triangles represent glycines.

## 4.10. Crystal packing

The unit cell consists of 6 asymmetric units, each of which contain two KorB-O—$O_B$ complexes. In total the unit cell consists of 12 such complexes (Fig. 16). The DNA duplexes stack on top of each other and form distorted pseudo-continuous helices (Fig. 16 and 17).

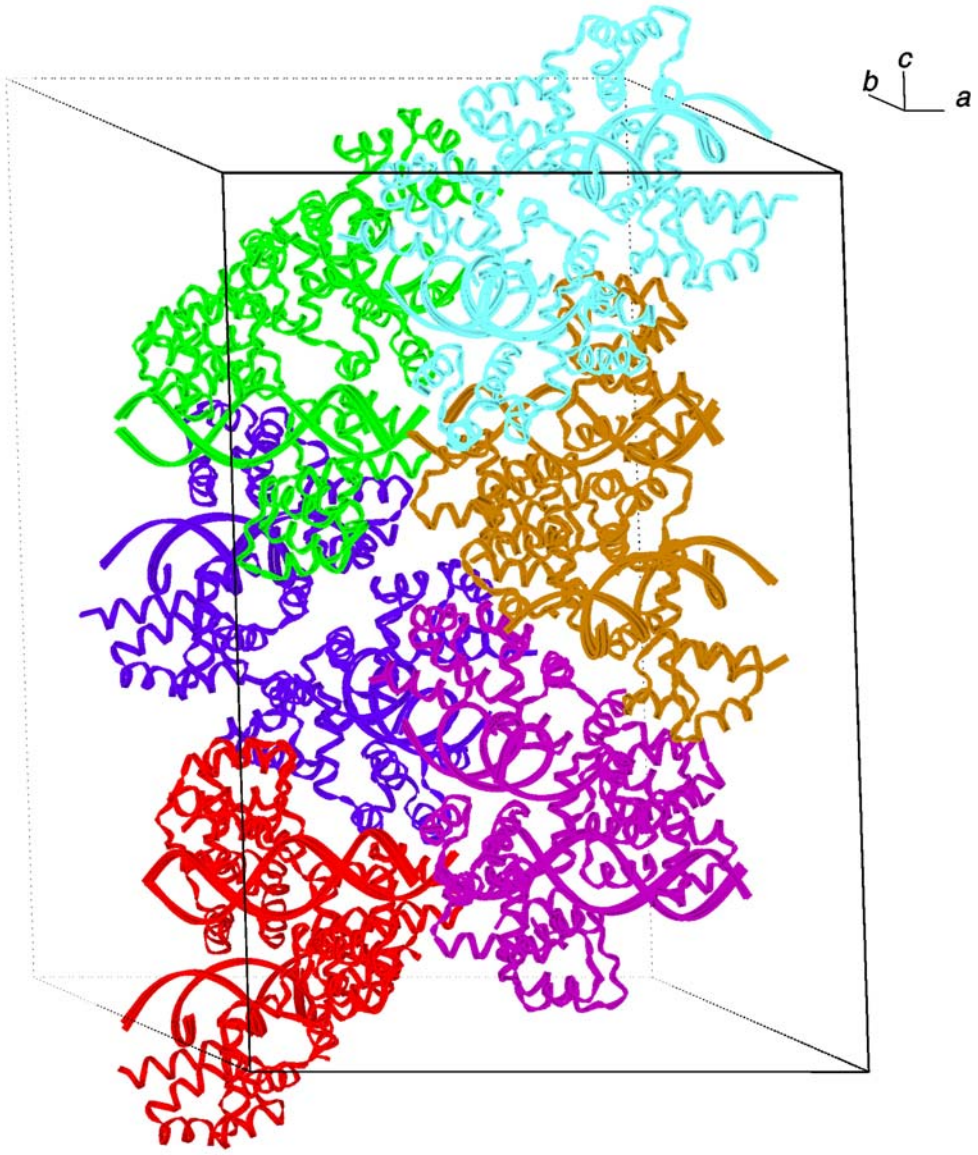



**Figure 16. Crystal packing of KorB-O—$O_B$ complex.** The figure shows an entire unit cell (space group $P3_221$) with 12 KorB-O—$O_B$ complexes. The DNA duplexes are stacked on top of each other and form pseudo-continuous helices.

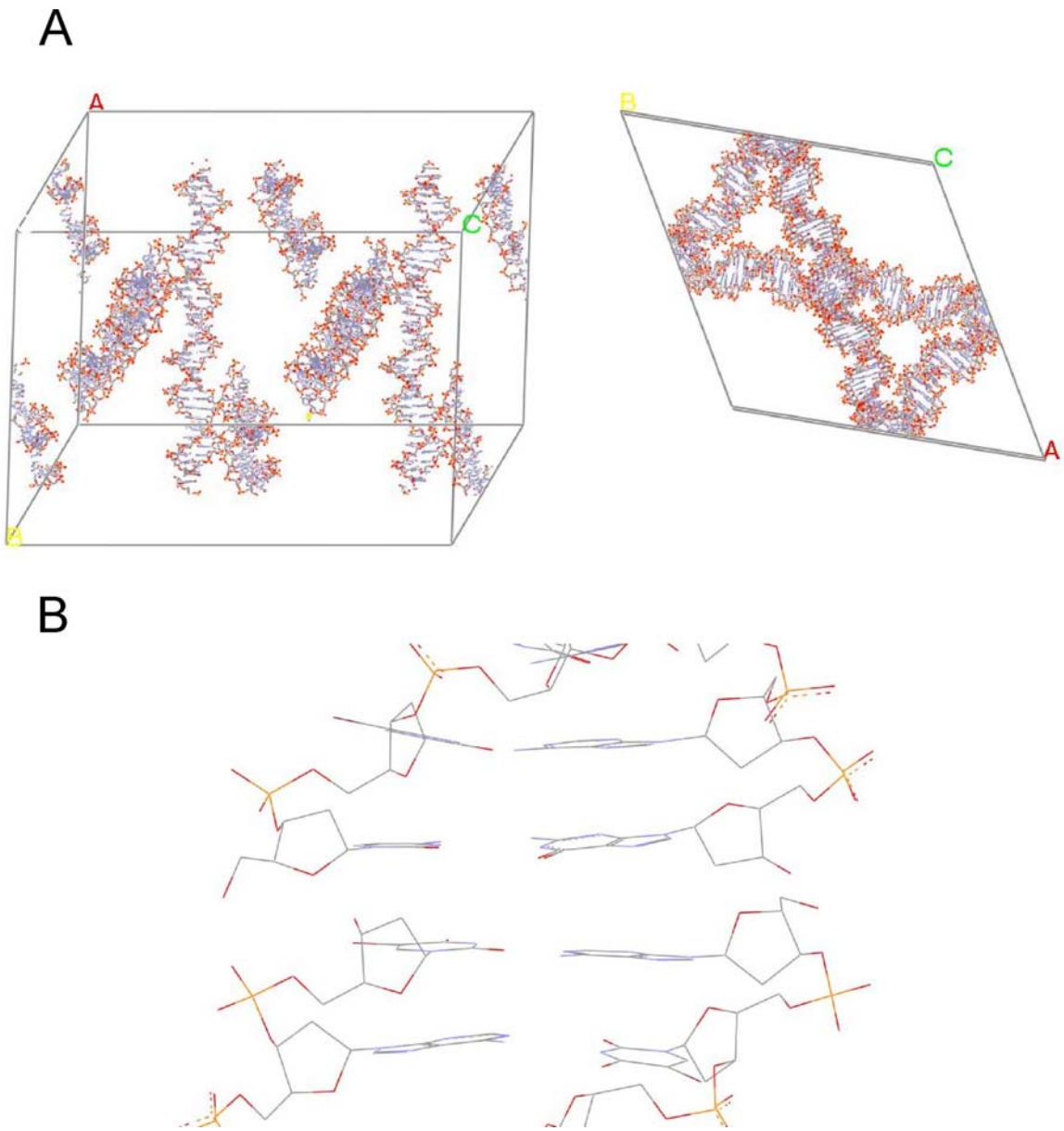



**Figure 17. DNA stacking in the crystal. (A)** An entire unit cell is shown in two different orientations but only with DNA molecules. (**B**) One duplex is packed symmetry-related on other duplex, forming a pseudo-continuous helix.

## 4.11. Structure analysis of the KorB-O—operator complex

The X-ray structure of the DNA-binding domain (KorB-O) of the transcriptional repressor and ParB homolog KorB was determined in complex with a 17-bp oligonucleotide containing the operator ($O_B$) sequence (Fig. 18).

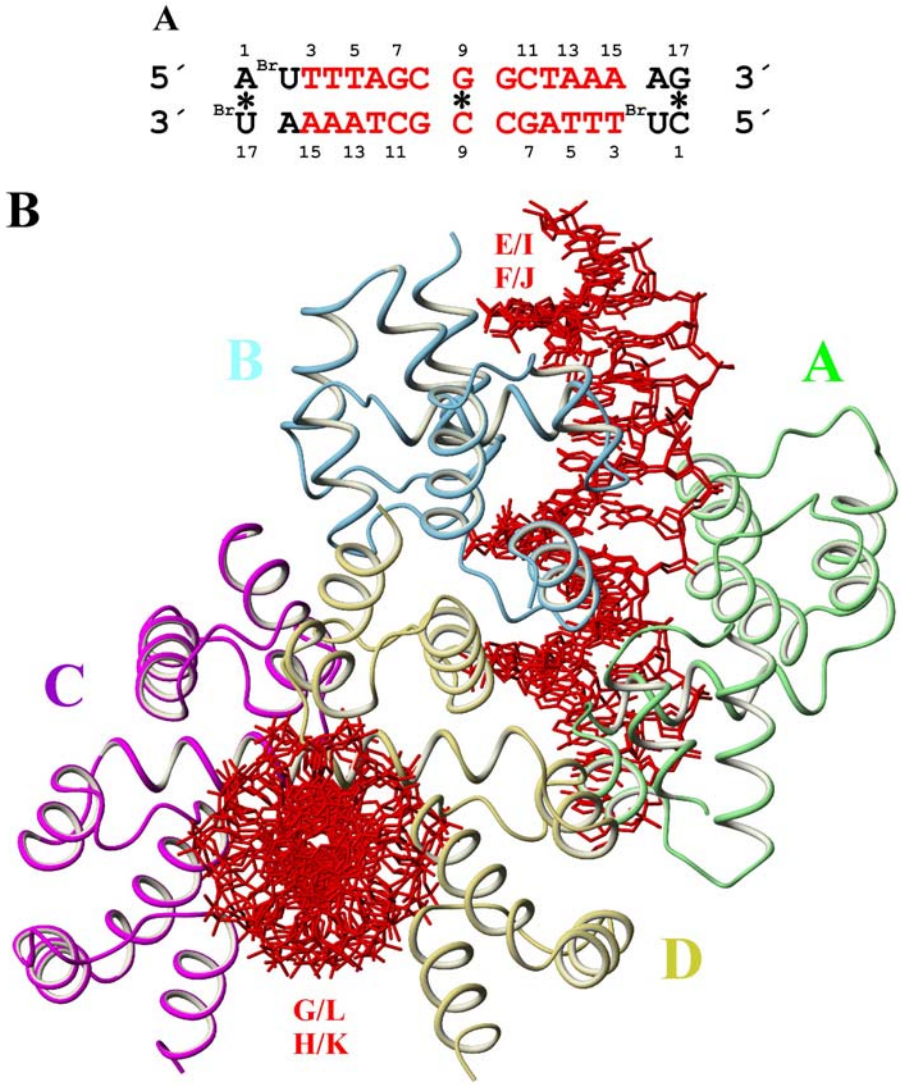



**Figure 18. Organization of the KorB-O—$O_B$ complex**. (**A**) Oligonucleotides used for co-crystallization with KorB-O. Base pairs written in black are outside the $O_B$ consensus sequence. Base pairs marked * break the two-fold symmetry of the DNA duplex. $^{Br}U$, 5-bromodeoxyuridine. (**B**) Asymmetric unit of the KorB-O—$O_B$ crystal. Protein strands A and B occupy the half sites in one operator complex, DNA-bound strands C and D comprise the other complex. The operator DNA is bound to KorB-O in two orientations where the KorB-O dimer AB binds the superimposed double strands EF and IJ, and KorB-O CD binds duplexes GH and LK. See text for details.

The protein and DNA fragments present in the complex are shown in Fig. 18. The multi-wavelength anomalous diffraction (MAD) phasing (Hendrickson, 1991) of the structure amplitudes was based on the anomalous scattering from bromine atoms incorporated into the DNA as 5-bromodeoxyuridine ($^{Br}$U) by chemical synthesis. Oligonucleotides were designed so as to carry the bromine label at base pairs outside the consensus $O_B$ sequence (Fig. 18A) in order to exclude any potential disturbance of protein binding by the bromines. The DNA fragment obeys perfect dyad symmetry for 14 base pairs which is only broken by the central G/C and the terminal A/$^{Br}$U or G/C base pairs. Iterative model building and crystallographic refinement converged at $R_{work}$ and $R_{free}$ of 19.5% and 25.0%, respectively (Table 5).

The final model, refined to a resolution of 2.2 Å, contains two complexes in the asymmetric unit; each complex is made up of two KorB-O molecules symmetrically bound to two operator half-sites (Fig. 18B). Although the KorB-O fragment used for co-crystallization with DNA covered the peptide region Arg117-Lys294, a significantly smaller domain, Tyr137-Glu252, could be modeled into the density. The missing N- and C-terminal residues appear to be disordered in the crystal, since an intact KorB-O fragment Arg117-Lys294 could be recovered from a dissolved crystal as proven by mass spectrometry and N-terminal sequencing (Appendix I and II). If these residues adopt a fixed conformation in intact KorB remains to be determined. As about 63 N-and C-terminal residues are missing per protein molecule, a total of 255 amino acids out of 712, approximately 35% of the protein scattering mass remains unmodeled.

Figure 19 depicts the secondary structural features of KorB-O along with an alignment showing the similarity between the full length KorB with four chromosomal members of the ParB family.

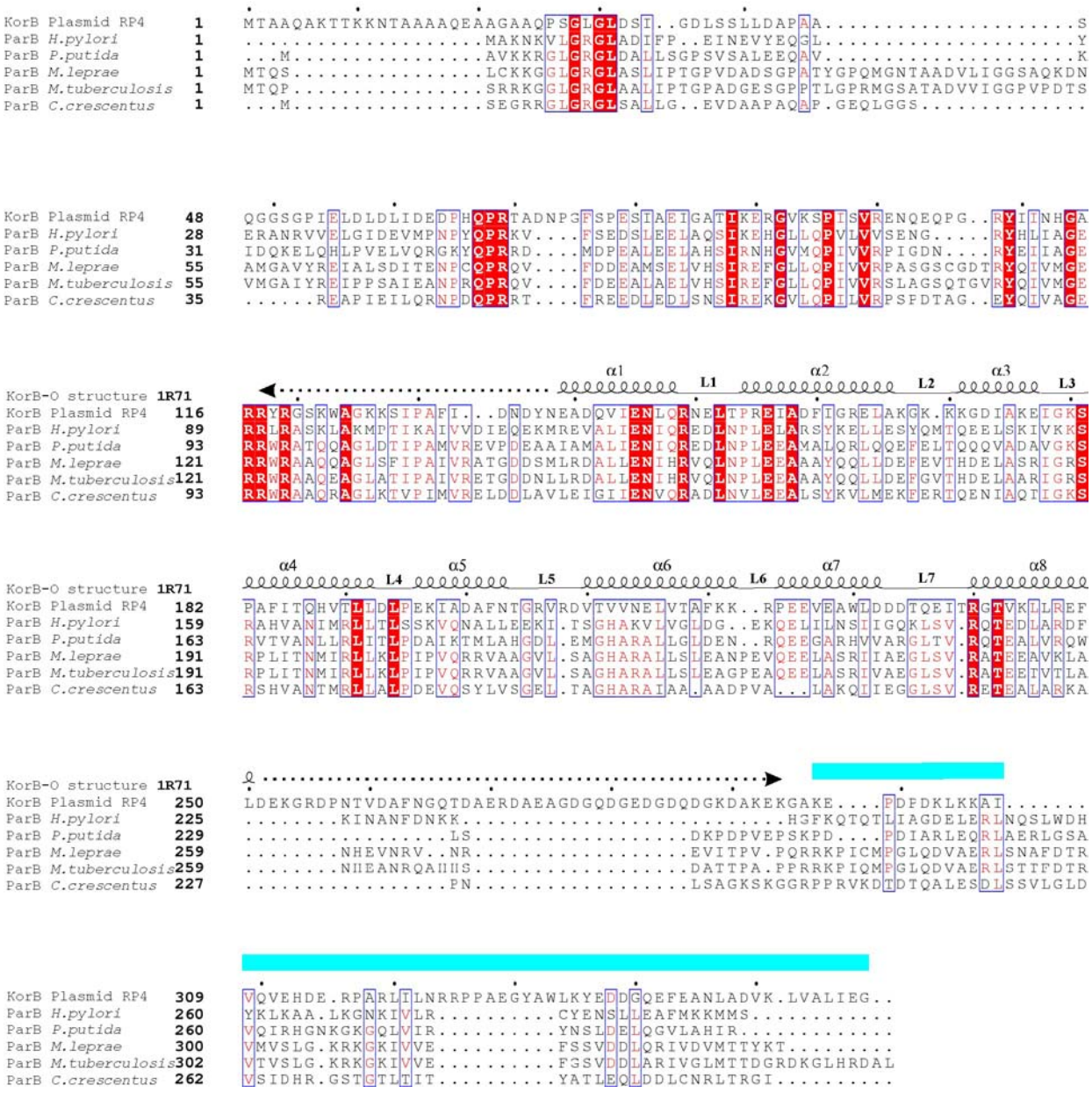



**Figure 19: Alignment of chromosomal members of ParB family along with KorB.** The secondary structure elements from the KorB-O structure (PDB code 1R71) are shown with helices (α1 to α8) connected by loops (L1 to L7). The cyan rectangle at the C-terminus is the KorB-C domain. Amino acid residues are highlighted and boxed in blue based on identity if it is more than 60 % in a column. Arrows mark the disordered regions in the crystal structure of KorB-O — $O_B$ complex. The sequence alignment was done with T-Coffee (Notredame *et al*., 2000) and the above output was generated with ESPript (Gouet *et al*., 1999). Swiss-Prot accession numbers for the proteins are KorB from plasmid RP4 (P07674), *Helicobacter pylori* ParB (O25758), *Pseudomonas putida* ParB (P31857), *Mycobacterium leprae* ParB (Q50201), *Mycobacterium tuberculosis* ParB (O53595) and *Caulobacter crescentus* ParB (O05190).

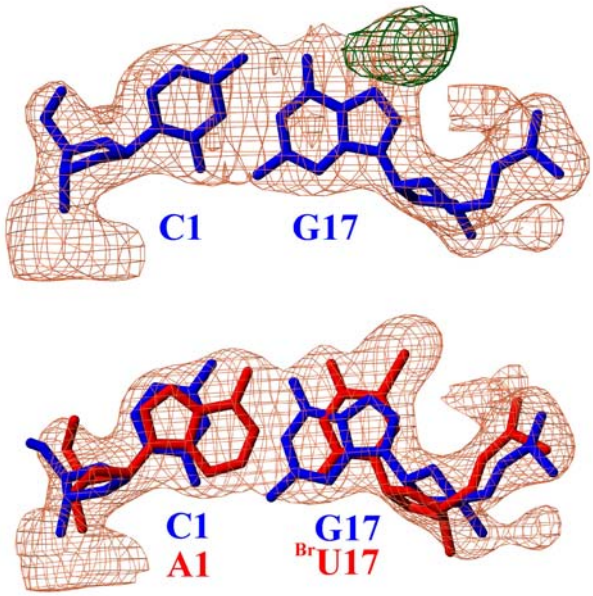



**Figure 20. Detection of disorder in the KorB-O bound operator DNA**. Top: Base pair C1-G17 with electron density before allowing for twofold disorder of the protein-bound DNA. Bottom: Base pairs C1-G17 and A1-$^{Br}$U17 after refinement of a model in which the DNA duplexes are superimposed in up and down orientation, both with 50% occupancy. The $2F_o$-$F_c$ density (coral) is contoured at 1.4σ, the $F_o$-$F_c$ difference density (green) is contoured at 4.5σ.

The final $2F_o$-$F_c$ electron density is well defined for the DNA (Fig. 20), if one allows for a two-fold disorder of the double helix which is made possible by the semi-symmetric sequence and results for each complex in an "up" and a superimposed "down" orientation of the duplex, each with 50% occupancy.

The final model (Fig. 18B), comprising four KorB-O molecules (strands A-D) and eight DNA single strands (E-L) displays good geometry with root-mean-square (rms) deviations of 0.015 Å and 2.012° from target values for bond lengths and angles, respectively (Table 5).

## 4.12. KorB-O consists of eight α-helices

The four protein chains present in the asymmetric unit of the crystal are represented by electron density to a different extent: Molecule A covers residues 139-252, B 137-252, C 140-251, and D 138-252, each bound to an operator half-site on the 17-bp oligonucleotide. The KorB-O monomer is composed of eight α-helices that are connected by loops (Fig. 21). The monomer can be subdivided into two subdomains: an N-terminal subdomain with the sequence-predicted HTH motif and a C-terminal up-and-down four-helix-bundle domain. The N-terminal subdomain contains four helices, α1 (residues 139-148), α2 (154-166), α3 (171-177), and α4 (182-192). Helices 3 and 4 are connected by a four-residue turn, and contain the residues predicted to form a HTH motif. The sharp turns between helices 2 and 3 and between helices 3 and 4 are made possible by Gly168 and Gly179, respectively, both adopting backbone conformations unfavourable for other residues in or near the shallow energy minimum of left-handed helix. Both glycines are conserved in all known KorB homologs. A short linker of four amino acids connects the N-terminal subdomain to the C-terminal subdomain consisting of helices, α5 (residues 196-203), α6 (210-222), α7 (224-232) and α8 (240-250). Helices 5, 6, 7 and 8 form a helical bundle and harbour the residues imparting specificity on operator binding.

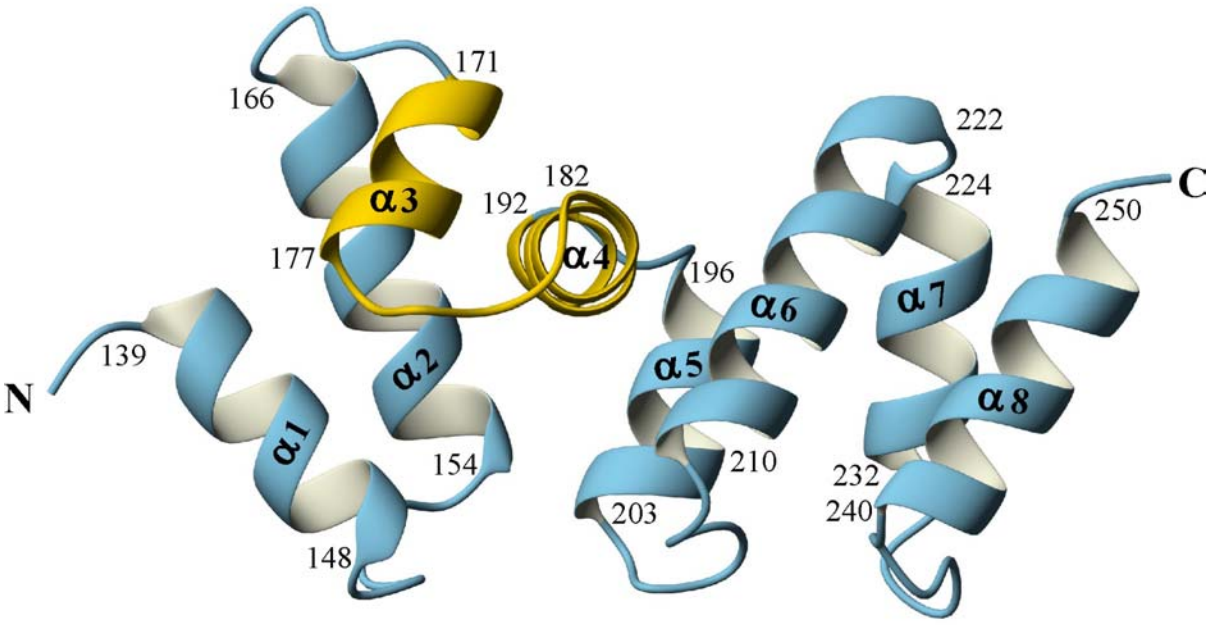



**Figure 21. Molecular structure of KorB-O**. α-Helices 3 and 4 (gold) represent the helix-turn-helix motif.

Two hydrophobic pockets are discernible in the KorB-O monomer structure. One involves helices α2, α3 and α4 and the conserved hydrophobic residues Ile161, Ile174, Ile178, Ile185, and Val189, as well as the non-conserved residues Leu165 and Phe184. This hydrophobic core serves to stabilize the conformation of the HTH motif and to anchor it in the KorB structure. The four-helix bundle of the C-terminal subdomain forms a second hydrophobic core with tight packing of hydrophobic side chains between the helices.
Elevated structural plasticity is seen at the N- and C-terminus of the KorB-O fragment as judged by the crystallographic *B* values (Fig. 22).

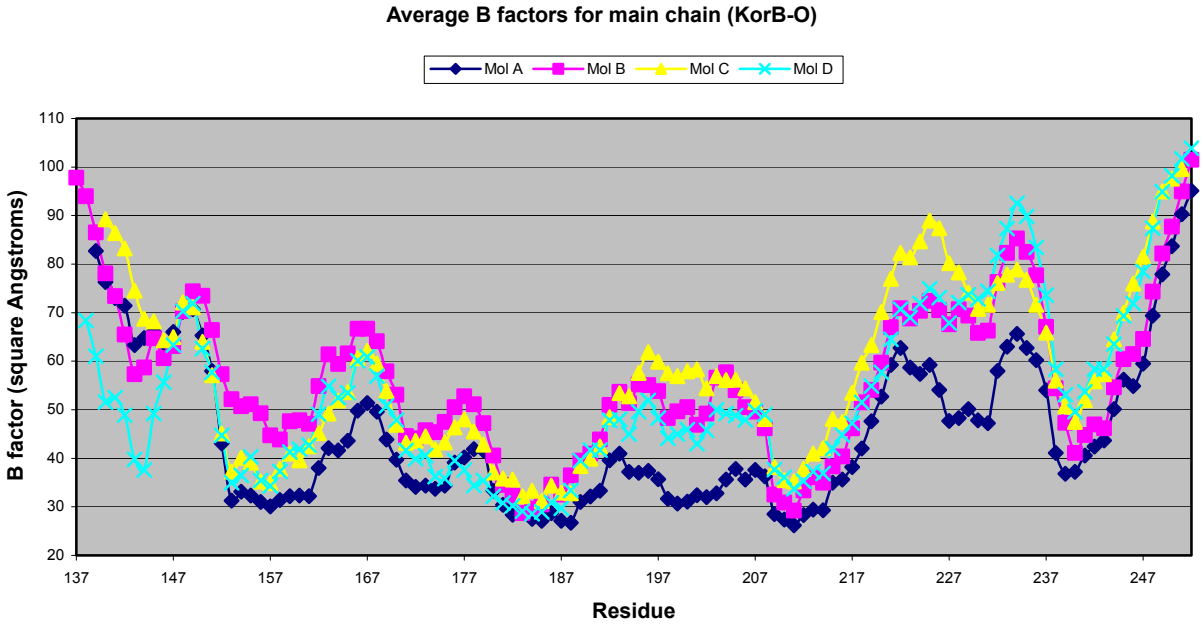



**Figure 22. Crystallographic *B* values for the main chain atoms of KorB-O molecule.** In order to compare the thermal motion in the four different molecules (A, B, C and D), a plot of *B* values against the residue number is shown.

Although unrelated by crystallographic symmetry, the KorB-O monomers show closely similar conformation, with rms deviations of 0.809 and 0.756 Å in a pair-wise least-squares superposition of corresponding $C^{\alpha}$ atoms 140-251 of monomers A-B and C-D. Superimposition of $C^{\alpha}$ atoms of each DNA-bound "dimer" AB and CD results in an rms deviation of 0.530 Å. Superposition of the DNA duplexes EF-IJ and GH-LK gives an rms deviation of 1.041 and 0.895 Å, respectively.

## 4.13. KorB binds to $O_B$ present in standard B-form

The pseudo-symmetric DNA duplexes in the co-crystal with KorB-O are averaged over two orientations, as seen in other protein-DNA complexes (Ferré-D'Amaré *et al*., 1993; Gomis-Rüth *et al*., 1998). The superposition of a 17-bp fragment in the "up" orientation onto one oriented "down" matches 14 identical base pairs, leaving a GC superimposed on a CG in the center and a CG matched with a $A^{Br}U$ at both ends of the double helix. Only the 14 matching base pairs are involved in protein contacts. The DNA disorder was unequivocally proven by the occurrence of prominent difference density at a terminal CG base pair which arose from the missing bromine atom at half occupancy (Fig. 20).

In the crystal, the KorB-O bound operator DNA adopts a standard B-DNA conformation with a straight helix axis lacking any major kinks. The sugar puckers vary around $C^{2'}$-*endo* ($C^{3'}$-*exo* to $C^{1'}$-*exo*) with occasional excursions to $O^{1'}$-*endo* or even $C^{4'}$-*exo*. The exocyclic backbone torsion angles vary around their expected mean values. The mean helix twist in the four-modeled duplexes is 35.2°, corresponding to an average helix repeat of 10.23 bp/turn. The mean rise per base pair is 3.28 Å. Not unexpectedly (Heinemann *et al*., 1992), the G/C-rich center of the helix has a widened minor groove with an average width of 7.1 Å, whereas near the A/T-rich ends the minor groove is narrowed to an average separation between phosphate groups of 3.2 Å. As a consequence, the major groove is slightly widened about 4 to 5 base pairs inwards from both ends where the "recognition helix" of the KorB-O HTH motif is inserted.

The presence of four crystallographically independent double strands in the crystals permits to determine local helix parameters with some confidence. A conformational analysis with the CURVES 5.0 program (Lavery and Sklenar, 1989) reveals that most global-axis parameters vary around mean values typical for B-DNA. A significant distortion, however, is observed for the symmetry-related base pairs T5-A13 / A6-T12 occurring twice within the $O_B$ consensus sequence of all four double helices. These base pairs show a prominent buckle with an average magnitude of 14.1 ± 3.0° which points in the same direction for adjacent base pairs. The base-pair step showing this prominent buckle marks a conformational shift from three base pairs, T3-A15 / T4-A14 / T5-A13 that are heavily propeller-twisted (21.6 ± 4.0°) to the two immediately adjacent base pairs A6-T12 / G7-C11 with essentially zero propeller, 0.6 ± 3.8°. This remarkable double-helical

conformation may be important for operator recognition by KorB: It will be shown below that one of the most important protein-DNA contacts addresses the C11 base.

## 4.14. Two KorB-O molecules bind symmetrically in the major groove of $O_B$

KorB-O binds in the major groove of the B-form DNA with each monomer contacting a half-site on the $O_B$ fragment (Fig. 23A). Four α-helices, two from each subdomain form direct contacts with the operator DNA. From the N-terminal subdomain, helices α3 and α4 make HTH-typical DNA contacts as observed in other bacterial repressor proteins. Helix α3 crosses the DNA major groove in perpendicular orientation. α4, corresponding to the "recognition helix" of other repressors, runs along the major groove, its axis only slightly offset from the plane of the DNA base pairs. Several amino-acid residues from these two helices make direct or water-mediated contacts with the DNA. Further amino acids involved in operator DNA binding are harboured in the helices α6 and α8 of the C-terminal subdomain. KorB-O contacts with the minor groove of the DNA are not observed.

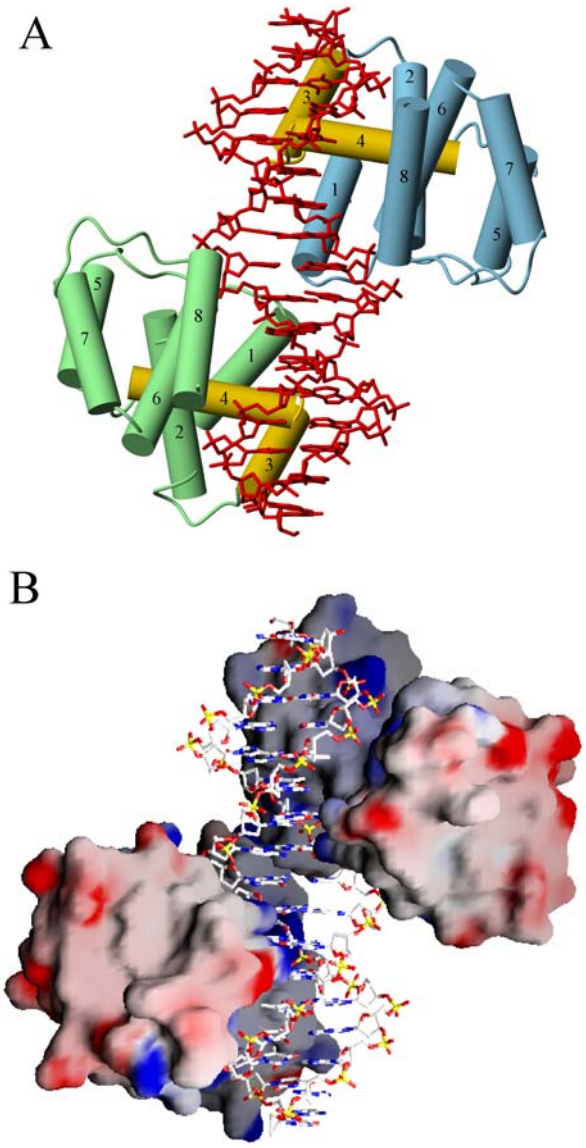



**Figure 23. The KorB-O—operator complex**. For the sake of clarity, the DNA is shown in only one of the two orientations present in the complex. (**A**) Schematic drawing of two KorB-O monomers occupying the half-sites of the $O_B$ operator site. α-Helices are depicted as cylinders. Helices forming the HTH motif of KorB-O are coloured yellow. (**B**) The electrostatic surface potential of KorB-O bound to $O_B$. Blue colour marks positive surface charge of the protein, red colour indicates negative charge.

A space-filling representation of the KorB-O—DNA complex (Fig. 23B) reveals that each protein subunit gently wraps around the DNA following the major-groove trajectory. Although KorB is an acidic protein with a calculated pI of 4.6 (Balzer *et al*., 1992), the surface contacting the DNA is positively charged throughout, probably contributing to the total binding energy through non-specific electrostatic attraction of the backbone

phosphate groups. Strikingly, the outward-facing surface of KorB-O is predominantly negatively charged.

From Fig. 24 it is also apparent that there is a protein-protein contact between the KorB subunits occupying the $O_B$ half-sites. This interaction extends from Glu145 to Glu151 and involves residues from the helix α1, and others in the turn between α1 and α2. KorB is known to be dimeric in solution (Balzer *et al*., 1992), and the C-terminal domain, KorB-C, has been identified as the main dimerization site (Delbrück *et al*., 2002). Whether the contact around Gln148 is important for KorB dimerization in solution in the absence of DNA remains to be determined.

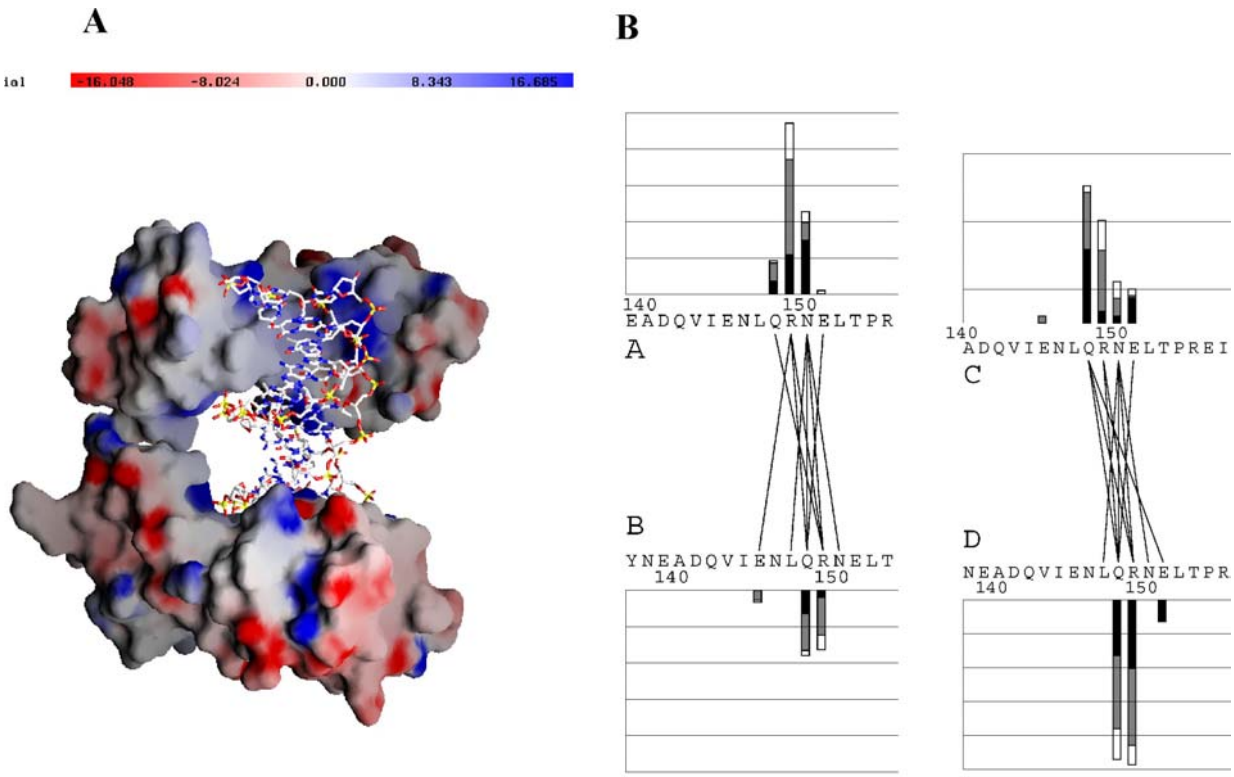



**Figure 24. Protein-protein contact between the KorB-O subunits**. (**A**) electrostatic surface potential of KorB-O bound to $O_B$. (**B**) XSAE linkage plot showing the interactions between the two subunits A - B and C – D from the two complexes in the asymmetric unit. Bars (scale 0 to 100) show how many $Å^2$ of each residue is occluded by the residue of the other chain. Polar contact areas are shown as black bars, hydrophobic contact area by open bars and mixed contact area between polar atom of one chain and hydrophobic atom of another chain as gray bars. Amino acids which occlude each other are connected by a straight line.

## 4.15. Exclusively Thr211 and Arg240 form hydrogen bonds to bases in $O_B$

The crystal structure of the KorB-O—operator complex reveals the determinants of sequence-specific $O_B$ binding very clearly. The half-site binding by KorB-O molecules described above and their interaction when bound to the DNA is essentially the same for KorB-O subunits A/B and C/D (Fig. 18B). Each individual protein-DNA contact is independently observed eight times in the crystal: In two KorB-O—DNA complexes, in two sub-sites per complex, and with the "up" and the "down" DNA duplex in each sub-site. The two-fold disorder of the DNA does not impair the protein binding, since KorB-O forms no contacts with the central base pair or the terminal base pairs of the 17-bp helix containing the operator sequence. In the following, we shall discuss only those inter-atomic contacts that occur at six or more out of the eight possible sites. This way we avoid possible confusion with contacts influenced by crystal packing or local structural fluctuations. Hydrogen bonds or salt bridges will have lengths of 3.5 Å or less, other contacts will be considered according to the Van-der-Waals radii of the atoms involved.

Only two KorB-O side chains are involved in direct hydrogen bonding to bases at each $O_B$ half-site. Thr211 accepts a hydrogen bond to its $O^{\gamma}$ from the $N^4$ of C11; Arg240 donates two hydrogen bonds from $N^{\eta 1}$ and $N^{\eta 2}$ to $O^6$ and $N^7$ of G10 (Fig. 25A). Both residues are anchored in the C-terminal subdomain of KorB-O, Thr211 at the N-terminus of helix α6 and Arg240 at the N-terminus of α8. Residues from the helices forming the HTH motif of KorB do not form direct base contacts in the $O_B$ region. The three hydrogen bonds to DNA from Thr211 and Arg240 are formed at each half-site. Hence, a total of six direct protein-DNA hydrogen bonds are formed at each $O_B$ operator. These interactions involve the inner GC base pairs of the 13-bp consensus operator sequence, but do not involve the central G9-C9 base pair. One of the participating base pairs, G7-C11, has an unusual geometry with zero propeller wherever it occurs, as described above.

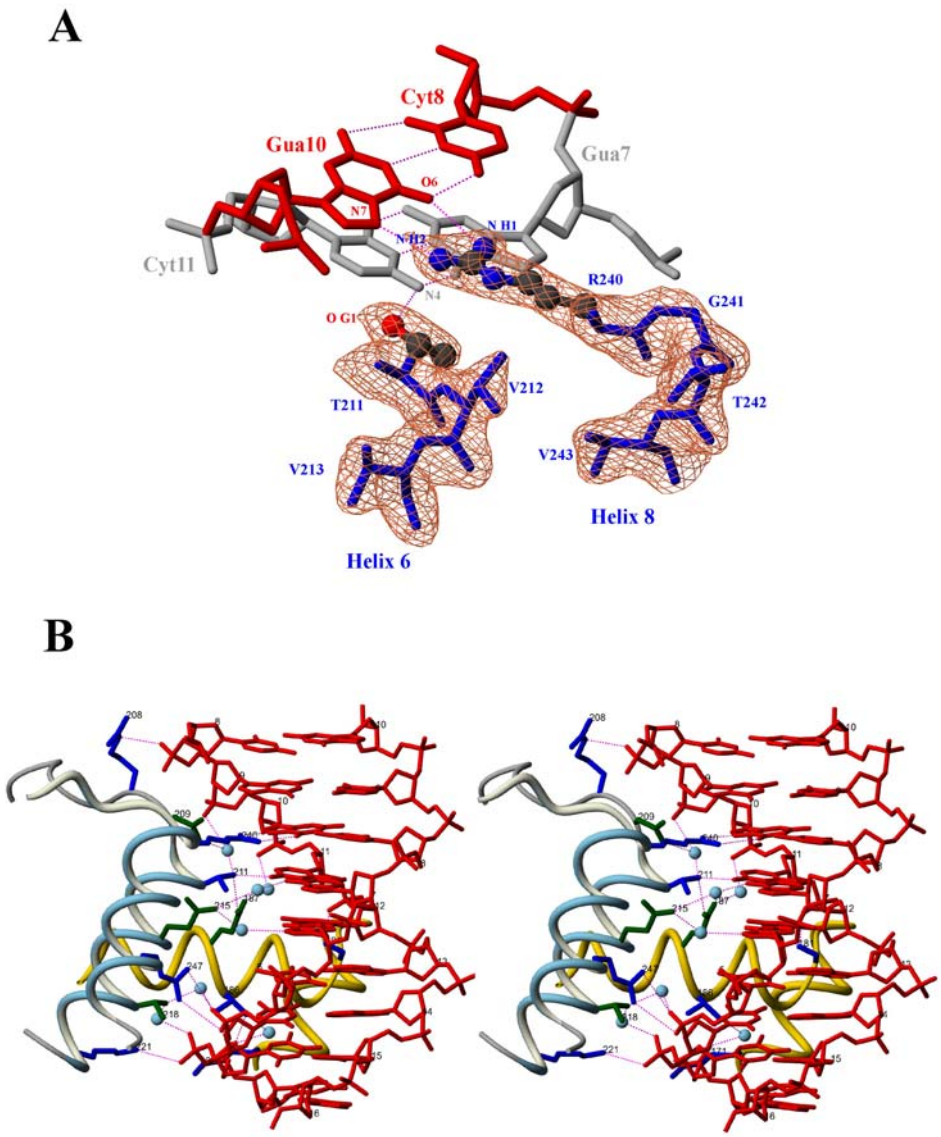



**Figure 25. Specific and non-specific KorB-O—$O_B$ contacts**. (**A**) Specificity-determining hydrogen bonding between the Thr211 (helix α6) and Arg240 (α8) side chains of KorB and C11 and G10, respectively, in one half-site of the KorB-O—operator complex. The $2F_o$-$F_c$ map around the protein region is contoured at 1.4σ. (**B**) Stereo view of the protein-DNA interface, water molecules are shown in sky blue, hydrogen bonding in magenta, residues involved only in water-mediated contacts are shown in green. Backbone-backbone interactions are not shown. The hydrogen bond involving the side chain of Arg208 is observed in only 4 out of 8 possible cases.

In addition to the small number of direct protein-base contacts, there are several hydrogen bonds and/or salt bridges linking KorB-O to the operator DNA *via* its sugar-phosphate backbone (Fig. 25B and 26). These contacts are formed by residues from the HTH motif (Lys171, Gly172, Ser181, Thr186) as well as amino acids from the C-terminal four-helix-bundle domain (Val210, Lys221, Arg247). Both Gly172 and Val210 are hydrogen-bonded through their backbone amide functions; in all other cases the protein side chain is involved. In total, there are 14 direct phosphate contacts per operator. In addition to the polar KorB-O—$O_B$ contacts discussed above, a number of Van-der-Waals interactions, mostly involving the methyl groups of thymines 3, 4 and 12 are observed. Residues from α-helices 1, 2, 5 and 7 do not interact with the DNA.

The KorB-O—operator complex is further stabilized by a number of water-mediated protein-DNA hydrogen bonds and Van-der-Waals interactions (Fig. 25B and 26). Essentially all DNA-facing residues of helix α4 are involved in water-mediated phosphate contacts, often through their backbone atoms. This reinforces the impression that this helix – the "recognition helix" in other bacterial repressors – plays a role in non-specific KorB-O—operator binding, but does not contribute to operator recognition. A single water-mediated contact addresses the DNA bases. Here, the carboxylate of Glu215 (α6) hydrogen bonds to two water molecules which in turn form hydrogen bonds to the base edges of A6 and G7. Again, this potentially specificity-conferring contact comes from a residue residing in the four-helix-bundle domain, outside the HTH motif.

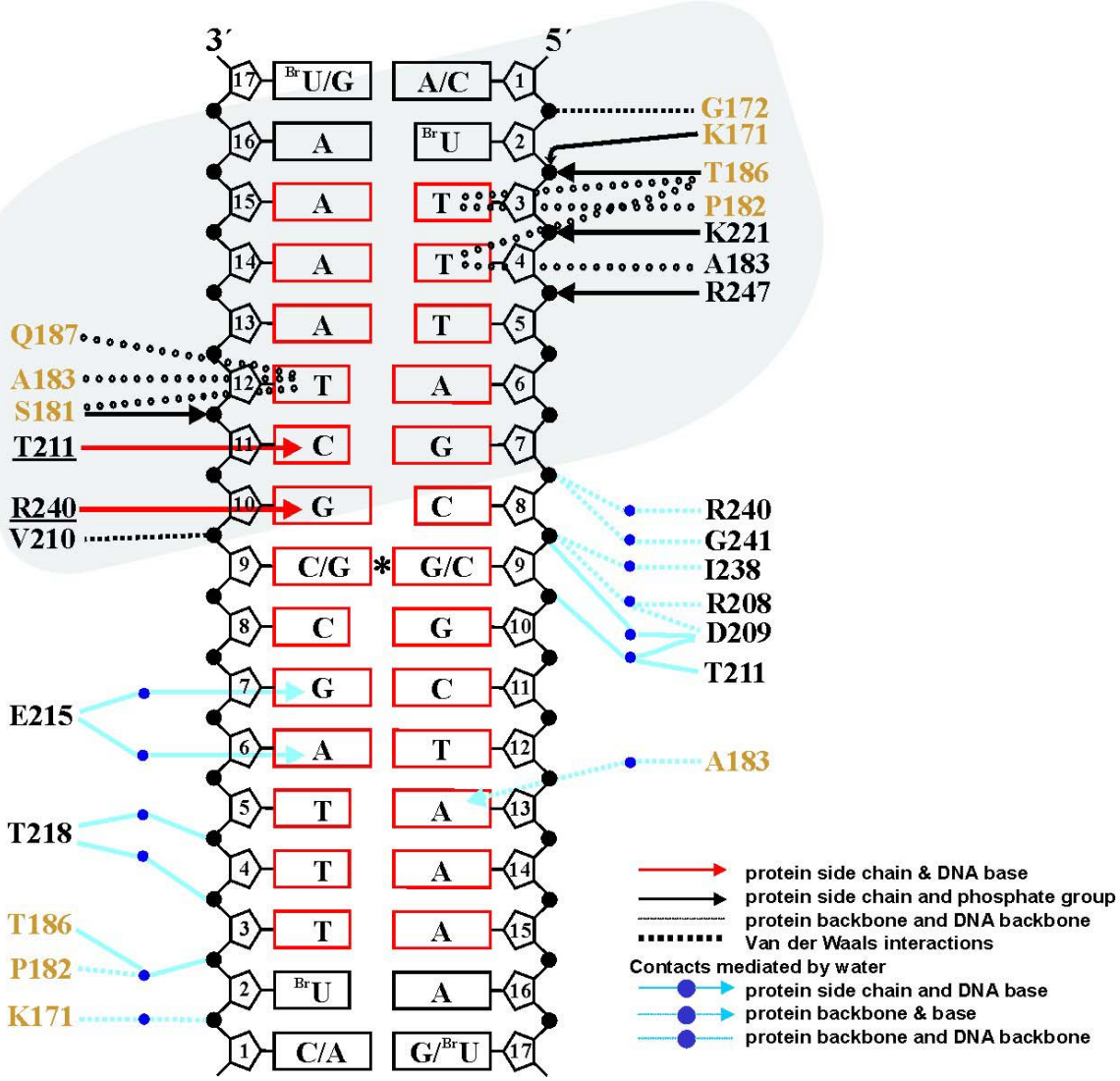



**Figure 26. KorB-O—DNA contacts**. The central 13-bp consensus $O_B$ sequence is shown in red, and the asterisk marks the two-fold symmetry axis. The symmetric operator has two half-sites, each of which binds one KorB-O subunit in identical geometry. Two different types of protein-DNA interactions are shown in the two half-sites. For the sake of clarity, in the top half-site (shaded) only direct protein-DNA contacts are shown, whereas water-mediated interactions are displayed in the bottom half-site. Both types of interactions are simultaneously present in both half-sites of the operator DNA. Amino acids numbered in gold are from the HTH motif, and underlined residues are specificity determinants. Direct hydrogen-bonded contacts involving protein side chains and DNA bases are depicted with red arrows, between protein side chains and phosphate groups with black arrows, and between protein backbone and DNA backbone with small dots. Van der Waals interactions are represented by large dots. Contacts mediated by water molecules (blue spheres) between protein side chains and DNA bases are marked with blue arrows, between protein backbone and base with a blue arrow with broken line, between protein side chains and DNA backbone with blue solid lines, and between protein backbone and DNA backbone with blue dotted lines.