Transcriptional Repressor CopR: Structure Model-Based Localization of the Deoxyribonucleic Acid Binding Motif

Katrin Steinmetzer,1 Alexander Hillisch,2 Joachim Behlke,3 and Sabine Brantl1*
1Institut für Molekularbiologie, Friedrich-Schiller-Universität Jena, Jena, Germany
2Institut für Molekular Biotechnologie, Jena, Germany
3Max-Delbrück-Zentrum für Molekulare Medizin, Berlin-Buch, Germany

ABSTRACT The plasmid pIP501 encoded transcriptional repressor CopR is one of the two regulators of plasmid copy number. CopR binds as a dimer to a nearly palindromic operator with the consensus sequence 5'-CGTG. Intermediate sequence searches revealed a significant structural relationship betweenCopR and the bacteriophage P22 c2 and the 434 c1 repressors. In this report we describe the experimental verification of a CopR homology model, which is based on a fairly low-sequence identity of 13.8 % to P22 c2 repressor. A model for the complex of CopR with the deoxyribonucleic acid (DNA) target was built on the basis of experimental footprinting data, the above-mentioned CopR homology model, and the crystal structure of the 434 c1 repressor-DNA complex. Site-directed mutagenesis was used to test the function of amino acids involved in sequence and nonsequence-specific DNA recognition and amino acids important for correct protein folding. CD measurements were performed to detect structural changes caused by the mutations. Exchanges of residues responsible for sequence-specific DNA recognition reduced binding to a nonspecific level. Mutations of amino acids involved in nonspecific DNA binding lead to decreased binding affinity while maintaining selectivity. Substitution of amino acids necessary for proper folding caused dramatic structural changes. The experimental data support the model of CopR as a helix-turn-helix protein belonging to the λ repressor superfamily.

Key words: plasmid pIP501; CopR; transcriptional repressor; HTH motif; protein-DNA interaction; CD measurements; site-directed mutagenesis; comparative modeling

INTRODUCTION

The binding of proteins to specific deoxyribonucleic acid (DNA) sequences is critical to the regulation of many cellular processes, among them replication, transcription, or recombination.

Protein-DNA interactions depend on specific tertiary configurations of the binding protein that allow the closest contact with the DNA helix. Therefore, they have been studied in detail over many years in both prokaryotes and eukaryotes. Although several prokaryotic transcriptional repressor proteins have been thoroughly characterized for DNA binding and oligomerization, e.g., Lac-repressor, λ-, Arc-, Trp-, or Tet-repressor, it is not so much known about transcriptional repressors involved in plasmid copy number control. Although it was published several years ago that CopB encoded by plasmid R1 is a tetramer,1 neither three-dimensional (3D) structures nor sequence motifs responsible for DNA recognition or oligomerization of this transcriptional repressor have been localized or characterized so far. However, for CopG encoded by the rolling circle-type plasmid pLS1,2 recently the 3D structures for both the protein and the protein-DNA complex were determined by X-ray crystallography. In contrast to preceding predictions, this very small protein (45 amino acids) binds to the DNA via a short β-sheet localized at the N-terminus, whereas the helix-turn-helix (HTH) motif was shown to be involved in dimerization.3 Recently, overexpression, purification, and gel filtration analysis of the pE194-encoded Cop protein has been reported.4 Plasmid pE194 belongs to the same class of rolling circle-type plasmids as pLS1,5 and its Cop protein is not expected to differ much from CopG encoded by pLS1.

The Cop proteins of the inc18 family of 0-replicating plasmids pIP501, pAMβ1, and pSM19035, CopR, CopF, and CopS, respectively, exhibit >95% sequence identity but are not related to CopG. CopR encoded by the broad host range streptococcal plasmid plp5016 is the best characterized of them so far. It is a 10.4-kD protein that is one of the two regulators of plasmid copy number7 and has a dual function: (a) it acts as a transcriptional repressor at the essential repR promoter by binding to a 44-bp inverted repeat upstream of and overlapping pII8 and (b) it increases indirectly transcription initiation from the antisense RNA promoter.9 CopR binds asymmetrically to two consecutive sites (I and II) within the major groove of the DNA and can dimerize in solution.10 Furthermore, CopR binds to DNA as a dimer and not as a monomer.11 The Kd value for CopR dimers was determined to be 1.49 ± 0.5 × 10⁻⁶ M, whereas the Kd, value for the dimer-DNA

Grant sponsor: Deutsche Forschungsgemeinschaft; Grant number: Br1552/4-1.
Alexander Hillisch’s present address is EnTec GmbH, D-07745 Jena, Germany.
*Correspondence to: Sabine Brantl, Institut für Molekularbiologie, Friedrich-Schiller-Universität Jena, Winzerlaer Str. 10, D-07745 Jena, Germany. E-mail: Sabine.Brantl@rz.uni-jena.de
Received 10 June 1999; Accepted 15 October 1999
complex was calculated to be $0.4 \times 10^{-10}$ M. The intracellular concentration of CopR in logarithmically grown *B. subtilis* cells was calculated to be $20–30 \times 10^{-6}$ M, suggesting that CopR also binds as a dimer in vivo.11

CopR has been subjected to a series of secondary and tertiary structure prediction methods.12 A 3D structural model of the first 63 amino acids of CopR binding as dimer to the DNA target was built. Based on this model a series of mutants was designed and assayed for DNA binding and the ability to dimerize to dissect the DNA recognition and dimerization motifs. Our data suggest that the predicted HTH motif localized between amino acid residues 18 and 37 is used for DNA binding. In addition, the function of one of eight amino acids (F5) predicted by the model to be involved in dimerization and the hydrophobic core could be confirmed experimentally.

**Model Building of CopR**

Based on the sequence similarity of CopR with the N-terminal domain of the structurally known P22 c2 repressor (PDB-code: 1ade)22 a homology model was built. The sequences of the structurally related proteins P22 c2 repressor (SWISSPROT accession number: P03035) and phage 434 c1 repressor (P16117) were aligned with the CopR sequence by using Clustal W.17 The homology model of CopR was built for the first 63 residues of the protein. The C-terminal part of the protein shows no sequence similarity to P22 c2 repressor and does not seem to be important for the biological function of CopR (see first paragraph of Results).

The model, including backbone, side chains, and the two amino acid deletion in CopR corresponding to residues 16 and 17 in the P22 c2 repressor was built by using the program MODELLER Version 4.22 The quality of the initial model was evaluated with the MODELLER energy command and WHATCHECK.23 The model was energy optimized by using the AMBER 91 force field24 within AMBER Version 4.1.25 A distance-dependent dielectric function of $\epsilon = 4r$ (where $r$ is the distance between two interacting atoms) was used to approximately simulate the surrounding aqueous solution. A residue based “twining-range” cutoff of 10/15 Å was applied for the treatment of nonbonded interactions.25 During the first 100 optimization steps a steepest descent minimization was performed. The energy optimization was then switched to the conjugate gradient method and terminated when the gradient norm of successive steps was below 0.1 kcal/(molÅ). This method has been shown to perturb high-resolution structures minimally, while improving the quality of a model.26 For comparison, the structure of the P22 c2 repressor (amino acids N2–K66) was subjected to the same minimization protocol. The RMS deviation of the nonminimized to the energy minimized backbone of the P22 c2 repressor is 0.3 Å, showing only minimal structural changes. The quality of the final model was evaluated with WHATCHECK25 and ProTable.27

The model of the CopR dimer bound to its 20-bp DNA target was built by fitting the backbone of the CopR monomer model onto the X-ray structure of the phage 434 repressor in complex with a 19-bp DNA fragment.28 This DNA target contains—as the CopR operator sequence—one consensus and one nonconsensus half site. The sequence of the DNA was replaced according to the protection pattern of the footprinting experiments of CopR with its target10 by using the SYBYL Biopolymer module.27 Because of an insertion of two residues (T43 and I44 in CopR) between P42 and R43 in the 434 repressor, a larger loop is observed in the dimerization interface. Thus, monomer I of CopR was slid along helix III in the major groove of the DNA to remove steric clashes of this additional loop and to build a proper dimerization interface. Flexible protein side chains at the surface of CopR were adjusted manually to accommodate the DNA and to remove unfavorable contacts in the dimerization interface. The same energy minimization protocol as described above

**MATERIALS AND METHODS**

**Sequence Alignment**

Sequence alignments and secondary structure predictions of CopR were performed during the FEBS Advanced Course “Frontiers of Protein Structure Prediction 1997” and are described only briefly here. Within the course, CopR was referred to as target no. 0237.12 The PDB90S sequence database of proteins with known 3D structure was searched for sequences related to CopR by using the BLAST method13 within SCOP.14 No related protein was found in this database. Therefore, a BLAST search (BLASTP, Version 1.4.9) with the sequence of CopR against the quasi nonredundant GenBank database at the NCBI (NRDB: 282,617 sequences) was performed to obtain related sequences. The BLOSUM62 comparison matrix15 was used. To mask segments of low compositional complexity in the query sequence, such as the acidic tail of CopR (residues E66–E86), the SEG filtering method16 was applied. No further related proteins were found in searching the NRDB with FASTA (see Ref. 35). A multiple sequence alignment of eight proteins that were found to be related to CopR with BLAST, was performed with Clustal W, Version 1.6.17 A gap opening penalty of 10.0 and a gap extension penalty of 0.05 were applied. The BLOSUM15 series of comparison matrices was used. In this program, the actual matrix used in an alignment step depends on the similarity of sequences that are to be aligned.17 Sequence identities and the amount of redundancy were calculated with BELVU Version 2.3.18 Intermediate sequence searches (ISS) were performed by comparing the sequences related to CopR against the PDB90S database with SCOP.14 ProDom Version 34.119 a database of protein domains, was also searched for sequences related to CopR. The HTH program20 was used to identify a possible HTH motif.
for the CopR monomer was applied to optimize the complex of the CopR-dimer with DNA.

**DNA Preparation and Manipulation**

Plasmid DNA was isolated from *B. subtilis* as reported previously. DNA manipulations (restriction enzyme cleavage, ligation, etc.) were performed at the conditions specified by the manufacturer or according to standard protocols. A GenAmp polymerase chain reaction (PCR) kit from PerkinElmer/Cetus was used for PCR. DNA sequencing was performed according to the dideoxy-chain termination method with a Sequenase kit from Amersham Pharmacia Biotech.

**Construction of E. coli/B. subtilis Shuttle Vectors Containing Mutated copR Genes and Determination of Plasmid Copy Numbers in B. subtilis**

Construction of plasmid pCOPΔ27 for the expression of a 3’- truncated copR gene was performed in two subsequent steps. First, two PCR fragments were generated on plasmid pUC119-F as template with either oligonucleotide B 198-30 (5’-GATTAATCCAAATGACCACTGACCACA) and the reverse sequencing primer or oligonucleotide B 199-30 (5’-TTGTTCTGTGGGCTAATTGGAATTAATC) and the universal sequencing primer and used as templates for a second PCR reaction with reverse and universal sequencing primer to obtain a 2.3-kb fragment. This fragment was cleaved with EcoRI and BamHI and inserted into EcoRI/BamHI digested pBT4 yielding plasmid pPR29Q. The obtained point mutation (nt 160 G-T) was confirmed by sequencing. Second, a functional copR gene was reconstituted by insertion of a KpnI/XbaI fragment from plasmid pCOPR29Q into the unique EcoRI site of plasmid pPR2. The obtained point mutation (nt 160 G-T) was confirmed by sequencing. Second, a functional copR gene was reconstituted by insertion of a KpnI/XbaI fragment from plasmid pCOPR29Q into KpnI/XbaI digested pBT4 resulting in plasmid pPRΔ27.

Plasmid pCOPLIC2 for the expression of pCOPK10Q was constructed as follows: First, two PCR fragments were generated on plasmid pCOP1B1 by using oligonucleotide 486-34 (5’-TCTAGAGGATCCGAAATTCGGATTTGTTTAAAAAATGGG) and oligonucleotide C 392-36 (5’-GAATATTAAGGAAGCTTACGAAAAAGTAGGAGGTACC) or oligonucleotide C 393-36 (5’-GGTACCTGACTTTTGGCTGAAGCTTCTCTCTAATG) and the reverse sequencing primer. Both PCR fragments were used as templates in a third PCR to obtain a 549-bp fragment that was cleaved with EcoRI and inserted into the unique EcoRI site of plasmid pPR1.

Plasmids pCOPS28T and pCOPE36G were constructed in the following way. First, two PCR fragments were made on plasmid pCOPS28T by using oligonucleotide 486-34 and oligonucleotides SB43 (5’-AAGAAATTGAATATGAATCTATAGAATTAT) or SB50 (5’-ATTACTGCAAGTTGATGATCGAATGAGAATCAGA), respectively, or the reverse sequencing primer and either oligonucleotide SB44 (5’-GAATATTAAGGAAGCTTACGAAAAAGTAGGAGGTACC) or oligonucleotide SB51 (5’-GGTACCTGACTTTTGGCTGAAGCTTCTCTCTAATG) and the reverse sequencing primer. Second, a third PCR reaction was performed with the PCR fragments 1 and 2 for the corresponding mutation and oligonucleotides 486-30 and the reverse sequencing primer. The resulting 549-bp fragment was cleaved with EcoRI and inserted into the EcoRI site of plasmid pPR1E derived from pPR1 carrying an erythromycin instead of a phleomycin resistance gene. The mutations and the fragment orientation were confirmed by sequencing.

The same approach as for pCOPS28T and pCOPE36G was used for all other mutant copR genes (mutagenic primers [pr1] and [pr2] are shown in brackets) resulting in plasmids: pCOPR29Q (pr1: 5’-GATTAATCCAAATGACCACTGACCACA) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) and pCOPR34Q (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) yielding plasmid pBT4 (pr1: 5’-AAGAAATTGAATATGAATCTATAGAATTAT) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) and pBT4 (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) yielding plasmid pBT4 (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) yielding plasmid pBT4 (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) resulting in plasmids: pCOPR29Q (pr1: 5’-GATTAATCCAAATGACCACTGACCACA) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) and pCOPR34Q (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) and pCOPR34Q (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) and pCOPR34Q (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) and pCOPR34Q (pr1: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC) (pr2: 5’-TTGTTCTGTGGGCTAATTGGAATTAATC).

**Construction of E. coli Vectors for Overexpression of Mutated copR Genes**

In all cases, a single PCR step was used to amplify the mutated copR genes from the corresponding *E. coli/B. subtilis* shuttle vectors. All amplified fragments were inserted into expression vector pQE9 (Quiagen). In this way, all mutated copR genes contain 11 additional 5’-codons encoding: Met-Arg-Gly-Ser-His-Gly-Ser fused to the second codon of copR (Glu). In the cases of pQED6, pQDR29Q, and pQRF3Q, PCR primers B618-30 (5’-GAATTCTGATTGCAGAATCTGTTTGG) and 951-30 (5’-GAATTCTGATTGCAGAATCTGTTTGG) were used, the PCR fragments were cleaved with BamHI and PsI and inserted into pQE9 digested with the same pair of enzymes. In all other cases, primers B618-30 and SB52 (5’-GAATTCTGATTGCAGAATCTGTTTGG) were used, the PCR fragments were cleaved with BamHI and inserted into the unique BamHI site of pQE9, giving rise to expression vectors pQS28T and pQE36G.

**Preparation of Labeled Wild-Type and Mutated CopR Targets**

DNA fragments used in the experiments are shown in Table I. 32P-labeled and unlabeled CopR targets were prepared as described previously. All oligonucleotides carry four G or C residues, respectively, at their ends to facilitate correct annealing and to promote additional stability.

**Preparation of Crude Cell Extracts, Purification of CopR, and Determination of Protein Concentration**

*B. subtilis* strain DB104 [pCOP7]7 DB104[pCOP20] or DB104[pCOPΔ27] expressing wild-type or 3’-terminally truncated copR, respectively, were grown on TY with 1 μg/mL phleomycin. In the case of all other mutated copR genes, DB104 strains with the corresponding plasmids were grown on TY with 5 μg/mL erythromycin. Crude cell extracts from overnight cultures were prepared by sonication as described.7 The supernatant obtained by centrifugation of the sonicated cells was stored at −20°C in 50% glycerol. Overexpression and purification of CopR from *E. coli* were performed as described before.11 The protein...
concentrations were determined by Bradford assay based on a calibration curve measured with His$_6$-CopR. An aliquot of the protein preparation used to obtain the calibration curve was subjected to amino acid hydrolysis to determine the concentration. For some of the proteins the concentration was also determined spectrophotometrically in the presence of 6 M guanidine-HCl by using the extinction coefficient of 1,280 M$^{-1}$ cm$^{-1}$ at 280 nm. Comparison of the values obtained by spectrophotometry with the Bradford-based values showed a fairly good agreement within an error range of about 10% for the proteins of wild-type length.

**CopR-DNA Binding Reaction and EMSA**

CopR-DNA binding reactions and electrophoretic mobility shift assay (EMSA) were performed as described previously.\textsuperscript{11}

**Cross-Linking**

Protein cross-linking with high-performance liquid chromatography (HPLC)-purified proteins was performed in 50 mM phosphate buffer pH 7.9 containing 150 mM NaCl and 0.0014% glutardialdehyde. The final protein concentration in the reaction mixture was 9 $\mu$M. Protein, buffer, and glutardialdehyde were mixed on ice and incubated at 7°C for 1 h. Increasing reaction time up to 4 h or increasing glutardialdehyde concentration up to 0.035 % did not increase the amount of cross-linked protein. Respective control experiments with $\alpha$-lactalbumin (which exists exclusively as a monomer) at a glutardialdehyde concentration of 0.0014% and incubation at 30°C for 1 h yielded dimers, trimers, and higher oligomers. In contrast, after incubation at 7°C, no such oligomers were observed (data not shown). Cross-linking of CopR at 30°C resulted in a higher amount of cross-linking products than at 7°C, but there was still a significant amount of dimers observable at the lower temperature, which was sufficient for our analyses. On the basis of these observations we were confident that the cross-linked dimers found under the conditions described above reflect only preformed dimers and not randomly cross-linked monomers. A glutardialdehyde-free control was always coincubated. After addition of loading dye containing 2% sodium dodecyl sulfate (SDS), the samples were heated at 95°C for 5 min, cooled on ice, and loaded onto a 17.5 % SDS polyacrylamide gel. Protein bands were visualized by Coomassie staining till saturation and subsequent destaining. For quantification of the cross-linked products the gels were digitalized with a Stratagene EAGLE EYE scanner and analyzed with TINA-Pcbs 2.0 software. Each experiment was performed with four to five independent parallels.
Analytical Ultracentrifugation

The molecular mass of CopR and mutated proteins was analyzed by means of an analytical ultracentrifuge XL-A (Beckman, Palo Alto, CA) using the sedimentation equilibrium technique as described previously. The radial absorbance distribution curves measured at three different wavelengths were fitted simultaneously by using the program POLYMOL. The concentration-dependent apparent molecular mass values (M_{app}) were analyzed according to a monomer-dimer equilibrium. An M_{1} value of 11,900 Da for wild-type and mutated CopR was used.

Circular Dichroism (CD) Measurements and Guanidine-HCl Denaturation

Sample concentrations of 0.5 g/L in 50 mM phosphate buffer containing 150 mM NaCl and 50% glycerol were used. The CD spectra were measured at room temperature in the range from 190 to 260 nm or 240 nm with a JASCO model 710 spectropolarimeter at a scan speed of 50 nm/min or 0.2 or 0.5-nm resolution. The path length of the cells used was 0.1 mm. The spectra were recorded as an average of 10–20 scans. The appropriate buffer baseline spectra were subtracted from the protein spectra. To calculate the mean residue ellipticity, the residue concentration used was obtained by multiplying the molar protein concentrations with the number of residues (103 aa for the His-tagged proteins CopR, K10Q, S28T, R29Q, R34Q, and E36G; 83 aa for CopRΔ20).

Guanidine-HCl-induced unfolding of the proteins was monitored by CD changes at 222 nm. Protein samples were incubated at the various guanidine-HCl concentrations for at least 3 h at room temperature to reach equilibrium. Control measurements after 24 h incubation showed no changes in CD signals compared with samples incubated for 3 h. The values were obtained after an accumulation between 5 and 10 and corrected for the buffer signal. The apparent fraction of unfolded protein for each guanidine-HCl concentration was calculated from the CD-signal by using the following equation: F_{app} = (Y_{obs} - Y_{nat})/(Y_{unf} - Y_{nat}) where Y_{obs} is the measured CD signal, and Y_{nat} and Y_{unf} were estimated from the extrapolated pre- and post-transitional baselines. To calculate the thermodynamic parameters of unfolding, we assumed a two-state process of unfolding where the folded dimers (D) are converted to unfolded monomers (U).

\[ D_2 \approx 2U \]

The equilibrium constant K is calculated as follows:

\[ K = 2P_1[F_{app}^2/(1 - F_{app})] \]

(1)

where P_1 is the total concentration of protein monomers. The free energy of unfolding is assumed to be a linear function of the denaturant concentration:

\[ \Delta G^0 = \Delta G_{H2O}^{0} - m[guanidine-HCl] \]

(2)

\[ \Delta G^0 = -RT \ln K \]

(3)

where R is the gas constant and T represents the absolute temperature. Based on the above equations the relation between F_{app} and the denaturant concentration can be written as follows:

\[ F_{app} = \exp[-\Delta G_{H2O}^{0} + m[guanidine-HCl]/RT]/4P_1 \times [(1 + 8P_1/\exp[-\Delta G_{H2O}^{0} + m[guanidine-HCl]/RT])^{0.5} - 1]. \]

(4)

The free energy change in the absence of guanidine-HCl, \Delta G_{H2O}^{0}, and m were calculated by fitting the above equation to the experimentally obtained F_{app} values by non-linear least-squares analysis using GRAPHIT version 3.01 (Erithacus, England).

RESULTS

Deletion Mutants of CopR-Copy Number Regulation In Vivo and DNA Binding In Vitro

To determine which parts of the CopR protein are essential for its function in vivo, we constructed a series of mutated copR genes and assayed them for their ability to regulate the copy number of pIP501 derivatives in vivo and for their DNA binding and heterodimerization properties in vitro. As shown previously, the deletion of up to 20 amino acid residues from the C-terminus of CopR still did not impair CopR function in vivo or in vitro. The plIP501 derivative pCOPΔ20 replicated at wild-type copy number (5–10 copies/cell). The DNA binding curve of purified CopRΔ20 was identical to that of wild-type CopR, and CopRΔ20 was able to form heterodimers with wild-type CopR. These data suggest that the 20 C-terminal amino acids of CopR are neither required for DNA binding nor for dimerization. A further truncation at the C-terminal part yielded CopRΔ27, a mutant with residual activity in vivo of about 20%. It was still able to down-regulate plIP501 copy number from 50–100/cell (plasmid without copR gene) to 25/cell. In EMSAs with \textit{B. subtilis} crude extracts, CopRΔ27 was able to form heterodimers with wild-type CopR, indicating that the dimerization motif was not affected significantly (not shown). From these results we could conclude that DNA binding and dimerization motifs of CopR are located within the first 65 amino acids of the protein.

Sequence Alignments and Similarity Searches of CopR

CopR is the hitherto best characterized representative of three related transcriptional repressor proteins (CopR, CopS, and CopF) encoded by streptococcal plasmids pIP501, pSM19035, and pAMβ1. These three proteins differ only by insertion or deletion of two amino acids at their C termini. A BLAST similarity search revealed three additional proteins as being significantly related to CopR (smallest sum probability P value between 3 and 8 \times 10^{-6}) and two probably related proteins (P value between 2 and 5 \times 10^{-3}). These five proteins are as follows: Xre, a putative \textit{Bacillus subtilis} prophage PBSX repressor (SWISSPROT accession number: P23789); the \textit{Mun}I regulatory protein from \textit{Mycoplasma} \textit{species} (SWISSPROT accession number: P43640); a \textit{B. subtilis} prophage PBSX repressor (PIR accession number: I40037); YQAE, a hypo-
A theoretical transcriptional regulator in the SPO III-CWLA intergenic region (SWISSPROT accession number: P45902); and a hypothetical protein 3 from Clostridium perfringens (PIR accession number: I40868). An initial multiple sequence alignment of all eight proteins was performed, and proteins showing >85% redundancy (CopS, CopF, and I40037) or a significantly shorter sequence length (P43640) were removed from the alignment. The multiple sequence alignment of the four remaining proteins (CopR, P23789, P45902, and I40868) is shown in Figure 1A. No 3D information on any of these proteins is available so far. Searching the PDB sequence database (SCOP14) with CopR returned no significant hit (lowest P value of 0.76). Thus, the proposed similarity of CopR to the P22 c2 repressor could not be detected directly by using BLAST or FASTA. However, when applying the intermediate sequence search technique by searching the PDB90S sequence database with any of the related three sequences (P23789, P45902, and I40868), it was found that all three proteins are significantly related to the P22 c2 repressor and to a lesser extent also to the 434 c1 repressor, the 3D-structures of which are known21,34. In all three cases the P22 c2 repressor was scored at position 1 with a significant relationship, and the 434 c1 repressor occurred at the second position. The multiple sequence alignment of CopR, the P22 c2, and the 434 c1 repressor reveals a sequence identity of 13.8% and 13.6% for the first 63 amino acids (aa) of CopR with the P22 c2 and 434 repressor, respectively (see Fig. 1B). Searching the ProDom-data-base19 with the CopR sequence also revealed the P22 c2 repressor as being related to CopR.

The probability that CopR comprises an HTH motif was calculated to be 71% by using the HTH program.20 The predicted HTH motif is located between residues 16 and 37. The model of the first 63 aa of CopR binding as a dimer to the target DNA fragment is shown in Figure 2. In analogy to the P22 c2 repressor this model suggests that CopR is an HTH protein and explains the general property of the protein to bind to DNA at two consecutive major grooves as a dimer.11 Within residues 1–63, five α-helices, two of which are involved in DNA binding, build up the protein backbone (see Fig. 3). Helix I is located between amino acids 5 and 13. Helix II comprising amino acids 18–25 is supposed to be the stabilization helix, whereas helix III spanning amino acids 29–37 is suggested to be the recognition helix. Furthermore, amino acids 44–54 and 58–62 are predicted to form helices IV and V. Amino acids R29, S30, S33, and R34 in the recognition helix are supposed to be the residues contacting certain bases in the DNA in a sequence-specific manner. The DNA phosphate backbone is proposed to be contacted by residues K10, K18, K20, S28, N31, and S40 in a sequence-unspecific manner. E36 is in close contact with K10, R13, and K18 which, in the model, are close together in space and in contact with the DNA phosphate backbone. The hydrophobic core of the
protein is built by residues F5, L9, F21, L25, Y32, I35, P42, L47, I50, and L53. Residues E2, F5, I44, K45, L47, L58, V59, and L62 occur on the surface of the protein and could be involved in dimerization. However, because of the uncertainties in the sequence alignment, the real conformation of the fifth helix involving residues L58–I63 may actually differ from the model.

The DNA binding specificity for CopR was compared with the observed specificity of the P22 c2 repressor using the method of Suzuki.36 The canonical binding positions for an HTH protein are residues aa n,a a n 1,a a n 1 4 and a a n 1 5, corresponding to N32, V33, S36, and Q37 in the P22 c2 repressor and Q28, Q29, E32, and Q33 in the 434 repressor. The respective residues in CopR are R29, S30, S33, and R34. Thus, it was predicted solely from the protein structure that R29 would preferentially contact a G (or with less probability also A or T) in position C1 (the Crick strand referring to Suzuki36 and the top strand in our study). R34 would contact a G (or with less probability also A or T) at position C3 or alternatively a G at position W4 (the Watson strand,36 our bottom strand). The predicted sequence for the DNA binding position of CopR consequently is 5'-GNGC-3'. The experimentally observed sequence of DNA binding site I is 5'-CGTGTC-3' (top strand, C239–T243, Table I, lower part) and of site II the sequence is 5'-CGTGGC-3' (bottom strand, starting with C255). Thus, site II contains the predicted recognition sequence 5’-GNGC-3’ (starting with G254–C251) whereas in site I the respective C/G pair at position 243 is exchanged by a T/A pair. Because the contact of an arginine to an adenine on the DNA is generally weaker than to a guanine, it was predicted that a modified target DNA with the palindromic consensus sequence 5’-CGTGTC-3' (C239–C243) in both sites should be an even better target than the wild-type DNA. Experimental data showed that this modified target is bound with at least the same high affinity as the wild-type DNA. On the other hand, a modified target containing a G251A substitution (KS3) lacking the predicted consensus G36 is bound with significantly decreased affinity.10 These observations support our model of CopR in which R34 should contact the respective G251 (Table I, lower part).

R29 of each monomer in CopR is predicted to contact the two outer G (G240 in binding site I and G254 in binding site II), respectively. R34 is proposed to contact A243 and/or G242 in binding site I and G251 (binding site II). Furthermore, the model suggests that E19 could make
water-mediated contacts with C239 and C255 explaining our footprinting data.\textsuperscript{10}

Localization of the DNA Binding Domain of CopR: Amino Acids 29–37 Constitute the Putative DNA Recognition Helix

Residues R29, S30, S33, and R34 within helix III are predicted to be the residues that recognize the DNA in a sequence specific manner.

Mutated proteins R29Q, R34Q, and E36G were constructed to verify the predicted function of helix III as the recognition helix. Moreover, the sequence alignments shown in Figure 1A and B indicate that glutamic acid in position 36 is conserved in all six proteins. Two other mutated proteins, K10Q and S28T, were constructed to prove the supposed nonspecific contacts of K10 and S28 to the DNA backbone (see above). K10Q was also intended to verify the function of E36. The model suggests that this glutamic acid holds the amino acids K10, R13, and K18 in place, three amino acids that contact the sugar phosphate backbone of the DNA nonspecifically (see Fig. 2). Therefore, both mutations E36G and K10Q should result in decreased binding affinity, but the operator selectivity should be preserved. Figure 3 shows a schematic representation of all constructed mutants.

A point mutation leading to an exchange of lysine at position 10 by glutamine (K10Q) resulted in a CopR mutant completely inactive in vivo. The copy number of the corresponding pIP501 derivative was 10- to 20-fold elevated compared with the wild-type (Table II) and as such comparable with the CopR deletion mutant. The same results were obtained for pIP501 derivatives encoding mutated proteins S28T and E36G (Table II).

According to the model where amino acid R29 is proposed to make specific contacts to DNA residues G240 (binding site I) and G254 (binding site II) and R34 is predicted to contact G242 or T243 (binding site I) and G251 (binding site II), it seemed to be interesting to analyze mutants R29Q and R34Q. Substitution of one of the critical arginine residues by glutamine should not only decrease the ability of CopR to recognize the wild-type target specifically, but might possibly enable these mutated proteins to bind a mutated target where the corresponding guanines in the DNA are substituted by adenines. Plasmids pCOPR29Q and pCOPR34Q replicated at approximately 10-fold elevated copy numbers in

<table>
<thead>
<tr>
<th>Protein</th>
<th>In vivo activity (%)</th>
<th>$K_D$ dimer (µM)</th>
<th>$K_D$ protein-DNA-complex (nM)</th>
<th>Relative amount of cross-linked dimer (%)</th>
<th>Activity of crude extracts from B. subtilis in EMSA (%)</th>
<th>$\Delta G_{H2O}$ (kcal/mol)</th>
<th>$m$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CopR</td>
<td>100</td>
<td>1.49 ± 0.49</td>
<td>0.4 ± 0.13</td>
<td>16.2 ± 1.6</td>
<td>100</td>
<td>10.07 ± 0.30</td>
<td>2.28 ± 0.16</td>
</tr>
<tr>
<td>CopΔ20</td>
<td>100</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td>100</td>
<td>8.7 ± 0.58</td>
<td>2.44 ± 0.44</td>
</tr>
<tr>
<td>K10Q</td>
<td>0</td>
<td>(10)\textsuperscript{a}</td>
<td>(17)\textsuperscript{a}</td>
<td>n.d.</td>
<td>11.5</td>
<td>9.88 ± 0.45</td>
<td>2.19 ± 0.24</td>
</tr>
<tr>
<td>S28T</td>
<td>0</td>
<td>n.d.</td>
<td>complex instable</td>
<td>n.d.</td>
<td>n.d.</td>
<td>9.69 ± 0.66</td>
<td>1.92 ± 0.34</td>
</tr>
<tr>
<td>R29Q</td>
<td>0</td>
<td>1.62 ± 0.2</td>
<td>no binding with NaCl</td>
<td>14.9 ± 1.3</td>
<td>6.55</td>
<td>not calculated</td>
<td></td>
</tr>
<tr>
<td>R34Q</td>
<td>0</td>
<td>1.88 ± 0.63</td>
<td>3450 ± 240</td>
<td>n.d.</td>
<td>4.3</td>
<td>8.9 ± 0.45</td>
<td>2.76 ± 0.39</td>
</tr>
<tr>
<td>E36G</td>
<td>0</td>
<td>(25)\textsuperscript{b}</td>
<td>(309 ± 13)\textsuperscript{c}</td>
<td>8.4 ± 1</td>
<td>18.3</td>
<td>not calculated</td>
<td></td>
</tr>
<tr>
<td>Dim6</td>
<td>0</td>
<td>(430 ± 178) \cdot 10\textsuperscript{a}</td>
<td>no binding</td>
<td>n.d.</td>
<td>0</td>
<td>not calculated</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}measured without NaCl
\textsuperscript{b}estimated from the crosslinking experiment
\textsuperscript{c}based on $K_D$ for protein-dimers estimated from crosslinking experiment

---

400 K. STEINMETZER ET AL.

Fig. 3. Amino acid sequence of CopR. Predicted $\alpha$-helices are shown as gray boxes and numbered above the boxes according to the text. Hatched boxes, amino acids predicted to be involved in dimerization. White boxes, mutated amino acids predicted to be involved directly or indirectly in DNA binding that were analyzed in this paper. Amino acid exchanges are indicated by arrows. The position of mutation CopRΔ20 is marked.
B. subtilis (Table II), indicating a loss of in vivo function of these mutants. The DNA binding properties of all mutated proteins were analyzed in vitro by EMSA by using B. subtilis crude extracts or purified proteins. Figure 4A shows the respective binding curves for the wild-type protein, R34Q and E36G obtained with purified His6-tagged proteins and the wild-type target KS1 in the presence of 75 mM NaCl. No DNA binding was observed for R29Q under these conditions but under salt-free conditions at micromolar concentrations, indicating nonspecific binding (data not shown). EMSA with S28T containing B. subtilis crude extract revealed that the protein-DNA complex was very unstable during the electrophoresis but could be stabilized by heterodimerization with CopRΔ13, a truncated mutant with wild-type activity (Fig. 4B). For the above-mentioned reasons, no binding constants and binding specificities were determined for S28T.

To test the binding specificity of the mutated proteins, we used several mutated DNA targets listed in Materials and Methods in Table I. The binding specificity was not affected in the case of K10Q and E36G, whereas R29Q and R34Q lost their binding specificity completely. This is in accordance with the CopR structure model.

To answer the question whether the weak DNA binding observed for the mutated proteins was solely due to the lack of a specific protein-DNA contact present in the wild-type protein-DNA interaction or also caused by alterations in the ability to dimerize, glutaraldehyde cross-linking experiments were conducted (Table II). Although the amount of cross-linked dimer for the wild-type protein and R29Q were comparable, fewer dimers were observed for E36G, indicating that the latter mutation affects dimerization either directly or indirectly. Furthermore, the Kd values for protein dimers of K10Q, R29Q, and R34Q were determined by analytical ultracentrifugation (Fig. 6 and Table II). Dimerization of R29Q and R34Q was basically unaffected by the mutations, suggesting that R29 and R34 were not involved in dimerization. In contrast, dimerization of K10Q was impaired compared with the wild-type protein CopR. The analytical ultracentrifugation measurements with K10Q were performed without NaCl. The value obtained for Kd was 40 × 10⁻⁶ M, 12-fold higher compared with the respective Kd value for wild-type CopR measured under the same conditions (3.3 × 10⁻⁶ M). On the basis of the Kd values for protein dimers and the obtained binding curves, we calculated the equilibrium dissociation constants for the protein-DNA complexes. The values are listed in Table II. Compared with
wild-type, DNA binding of K10Q was 15-fold (measured under salt-free conditions and compared with the respective value $K_d = 1.1$ nM for wild-type CopR), of R34Q 8,600-fold and of E36G 770-fold decreased, indicating that binding of R34Q is also reduced to a nonspecific level as observed for R29Q. Binding curves were analyzed as described in Steinmetzer et al.\textsuperscript{11}

Effects of the Alteration of One Amino Acid Predicted to Be Involved in Dimerization and Hydrophobic Core

Our model predicts eight amino acids involved in protein dimerization: E2, F5, I44, K45, L47, L58, V59, and L62, whereby F5 should contact I44 of the corresponding other CopR monomer (see Fig. 2). Furthermore, F5 is predicted to be part of the hydrophobic core. Based on this prediction a mutation of F5 should influence both dimerization and folding.

To verify the function of amino acid F5, shuttle vector pDim6 (F5V) and expression vector pQED6 were constructed. Mutated protein Dim6 was nonfunctional in vivo (Table II), and crude extracts from \textit{B. subtilis} containing pDim6 were, in contrast to crude extracts from all other mutants, completely inactive in EMSA experiments. Analytical ultracentrifugation showed that Dim6 is almost unable to dimerize in the concentration range used for the experiments (Fig. 6). The $K_d$ value was determined to be $430 \pm 178$ mM, indicating that this mutated protein is under physiological conditions (intracellular concentration of inactive mutants are $\sim 300$–$400$ nM) mainly present as monomers. Because the model predicts that F5 is involved in the hydrophobic core, the mutation F5V should

![Fig. 5. Binding specificity of wild-type and mutated proteins. A: EMSA was performed with radioactively labeled DNA fragments KS1, KS3, and KS5 (sequences listed in Table I) and following HPLC-purified proteins: wild-type (0.16 μM), R29Q (18 μM), R34Q (21 μM), and E36G (2.5 μM). The protein concentrations were adjusted for each protein to different values to yield detectable amounts of complex but avoid formation of the higher order complex (already visible on top of the R34Q-KS1 and R34Q-KS3 complexes). B: EMSA was performed with unlabeled DNA fragments and wild-type protein (upper panel, Reprinted from J Mol Biol 283 #3, Steinmetzer et al., pp. 595–603, 1998, by permission of the publisher Academic Press.\textsuperscript{11}) or K10Q (lower panel). The concentration of wild-type protein is 0.5 μM and of K10Q 2.4 μM. Complex and unbound DNA were visualized by ethidium bromide staining.](image-url)
result in a significant structural change, too. This was confirmed by CD measurements (see below and Fig. 7B).

Analysis of Structural Alterations and Protein Stability

Circular dichroism measurements were used to detect and evaluate structural alterations in the mutated proteins in comparison with the wild-type protein His6-CopR. The unusual buffer conditions (50 mM phosphate buffer, 150 mM NaCl, and 50% glycerol) used for monitoring the spectra were chosen because of the following observations: (a) the purified wild-type protein becomes inactivated within days when stored without 50% glycerol (at 4°C, −20°C and at −70°C) and (b) titrations of wild-type protein (stored in this buffer) with equimolar amounts of its DNA target showed that the protein is nearly 100% active under these conditions (data not shown). Thus, we assume that the observed CD spectrum for the wild-type protein reflects its biologically active conformation.

The far-UV CD spectra of wild-type and mutated proteins are displayed in Figure 7. The wild-type protein spectrum shows a positive band near 192 nm and two negative bands near 208 and 222 nm, indicating the presence of α-helical structures. A similar spectrum was obtained for S28T, indicating that the respective mutation did not cause structural alterations in the protein. The CD spectra of R29Q, R34Q, and CopR20 exhibit only slight differences compared with the wild-type spectrum. The CD spectrum of K10Q shows a slightly decreased α-helical content. In contrast to the above-described spectra, the CD spectra of E36G and Dim6 (F5V) look completely different: The α-helical content was strongly decreased, indicating that both single mutations, E36G and F5V (Dim6), caused structural changes in the protein conformation. Especially the spectrum of Dim6 resembled that of a partially unfolded protein.

To analyze the stability of the proteins, guanidine-HCl denaturation experiments were performed. To monitor the unfolding process, changes in ellipticity at 222 nm were recorded for the different guanidine-HCl concentrations. The transition curves are shown as insets in Figure 7. Monophasic cooperative transitions from the completely folded to the unfolded state were observed for the wild-type protein, S28T, K10Q with a midpoint of denaturation (c_m) at 1.8 M guanidine-HCl and for R34Q and CopA20 (Fig. 7A, inset). The latter mutant exhibits a decreased stability, because 50% unfolding is reached at 1.1 M guanidine-HCl. The transition monitored for R29Q is clearly biphasic with c_m1 = 0.95 and c_m2 = 2.6 M, suggesting the presence of an unknown intermediate state. A similar observation was made for one phosphoglycerate kinase mutant in contrast to monophasic transitions for the wild-type protein and several other mutants. However, the transition curves of E36G (not shown) and Dim6 differed from the above described in that at low guanidine-HCl concentrations the ellipticity at 222 nm decreases, which could be indicative for the occurrence of protein aggregation before the unfolding process proceeds at higher concentrations of denaturant. It has been shown that especially partially unfolded proteins have the propensity to aggregate.

In summary, according to the CD spectra the proteins can be divided into two groups: proteins with wild-type like or slightly changed structure (S28T, R29Q, R34Q, CopA20, and K10Q) and proteins with severe structural changes (E36G and Dim6).

Six additional mutated proteins were constructed to test the other amino acids predicted to be involved in dimerization (manuscript submitted). Indeed, all putative dimerization mutants revealed defects in the ability to dimerize as shown by increased K_D values of the corresponding dimers and were, thus, supporting the 3D model of CopR.

DISCUSSION

In this article we present a combination of structure modeling and mutational analysis to identify functional motifs of the transcriptional repressor CopR. Although the model was based on only low-sequence similarity between CopR and P22 c2 repressor, our subsequent experimental data verified important predictions of the presented model.

One important limitation of the model is that only the N-terminal 63 amino acids of CopR are considered. The model prediction of the tertiary structure of CopR yields 43 of 63 amino acids in α-helical conformation, which is consistent with secondary structure predictions of CopR. Therefore, considering 92 amino acids of full-length CopR, the α-helical content should be at least 47%, assuming that the last 29 amino acids do not form α-helices. Deconvolution of the measured CD spectra was performed with different programs. By using the program according to Yang et al., 39% α-helix, 25% turn and 36% random coil were calculated. However, deconvolution according to Manavalan and Johnson yielded 20% α-helix, 22% β-sheet, 24% turn and 32% other. The differences in α-helical and β-strand content between the calculations by Yang et al.
and Manavalan and Johnson are due to the different algorithms on which these programs are based. Because of the divergencies in our calculation results, the uncertainty in the prediction of α-helix V in our model (see Results), and the unknown structure of the C-terminus, we used the results of the CD measurements only to decide whether the introduced mutations caused structural alterations. In the case of CopG, calculations with different programs also yielded widely diverging results.41 Such divergencies can be resolved only by tertiary structural analyses as X-ray crystallography or NMR.

Exchanges of Residues Predicted to Be Responsible for Sequence-Specific DNA Recognition Reduced Binding to a Nonspecific Level

Generally, arginine residues preferentially recognize guanines on the DNA targets, whereas glutamine residues preferentially bind to adenines.35 This was the reason for the construction of mutants R29Q and R34Q.

DNA binding studies with both purified mutated proteins R29Q and R34Q showed that all mutated DNA targets (KS3, KS5, and KS7) that contain one mutated and one wild-type half-site, as well as the wild-type target KS1 were equally well recognized by R29Q, albeit only at very high protein concentrations (Kd values in the micromolar range). EMSA with R29Q homodimers and a double-mutated DNA target (G240A/G254A, data not shown) did also not prove the expected new binding specificity of R29Q. The other arginine substitution mutant, R34Q, also proved to bind only nonspecifically to the corresponding targets. Glutaraldehyde cross-linking and determination of Kd values for mutant dimers by analytical ultracentrifugation revealed that dimerization was unaffected in both R29Q and R34Q (Table II). CD measurements showed the same spectra for R29Q, R34Q, and CopRD20 (Δ). These spectra differed only slightly from the wild-type spectrum, indicating only small structural changes. However, because CopRD20 is 100% functional in vivo, the altered CD spectrum of this mutant—and also of R29Q and R34Q—characterizes only minor structural perturbations that might influence the stability of the proteins, but not their ability to bind or to dimerize. Indeed, guanidine-HCl denaturation experiments showed a decreased stability for R29Q, R34Q, and CopRD20. These experiments showed that the stabilizing effect of the mutations R29Q and R34Q was surprising because the side chains of both, R29 and R34, are expected to be solvent exposed and not involved in interactions with other amino acids side chains. But for several other nucleic acid binding proteins, such as the Trp repressor,42 Rop,43 and the Cro repressor of bacteriophage λ,44 it has been...

Fig. 7. Circular dichroism spectra and guanidine-HCl induced unfolding of wild-type and mutated proteins. Measurements were performed at 21°C at 0.5 g/L protein concentration. The insets show the fraction of unfolded protein plotted as a function of guanidine-HCl concentration. Unfolding data were obtained by monitoring CD signal at 222 nm. A: CD spectra of wild-type (——), K10Q (——), R29Q (——), R34Q (−−−−), S28T (−−−−), and CopRD20 (——). Inset: Unfolding curves of wild-type protein (•), K10Q (▲), S28T (▼), R29Q (▼), R34Q (□), and CopRD20 (Δ). B: CD spectra of Dim6 (——) and E36G (−−−−). For comparison the wild-type spectrum (——) was included. Inset: The unfolding curves of the mutated protein Dim6 (▲) and of the wild-type protein (▼) for comparison.
shown that mutations of surface located amino acids can cause both increasing or decreasing stability of the protein, whereas the overall structure is not or only slightly changed.

The following arguments might explain our failure to construct mutants with a new binding specificity. The prediction of the expected new binding specificity was based on the general recognition code proposed by Suzuki et al. But, even for the very closely related 434 c1 and 434 Cro repressors, a single amino acid exchange was not sufficient to create a new binding specificity. In the 434 c1 repressor, Q33 recognizes an AT pair, whereas L33 at the respective position in 434 Cro contacts a GC pair. Substitution of Q33 by L did not lead to a preference for GC in the 434 c1 repressor and vice versa. It is certainly true that sequence specific recognition of HTH-proteins is not solely mediated by the contacts of amino acids in the recognition helix but also by surrounding regions of the protein and exchanging only one amino acid at a time might not be sufficient to permit specific recognition of a new target sequence.

Mutations of Amino Acids Predicted to Be Involved in Nonspecific DNA Binding Led to Decreased Binding Affinity While Maintaining Selectivity

Amino acid S28 was predicted to contact the DNA backbone nonspecifically. CD measurements showed that mutated protein S28T has a wild-type-like structure and stability. Thus, we can assume that the observed decreased complex stability during gel electrophoresis is solely due to a lacking or impaired contact between S28 and the DNA, which is in agreement with the model. Amino acid K10 was also predicted to contact the DNA backbone nonspecifically. Binding affinity of K10Q was severely impaired, whereas binding specificity was not changed. K10Q shows certain structural alterations, compared with the wild-type protein, which are supposedly due to the proposed contacts to amino acid E36 (see below). Thus, analyses with both mutants S28T and K10Q also support the model.

Substitution of Amino Acids Necessary for the Proper Folding Caused Dramatic Structural Changes

For E36, the model predicts that this amino acid makes contacts to three other amino acids, among them K10. Therefore, it is not surprising that, as in the case of K10Q, the binding specificity of E36G was not changed, but the binding affinity was severely impaired. E36G revealed structural alterations compared with the wild-type protein, indicating that the interaction between E36 and K10 is of importance for the correct folding of the protein as to be expected from the model prediction. Apparently, E36G is a mutant with an altered structure, whereas K10Q is not. This observation is consistent with our model where E36 holds three other amino acids (K10, R13, and K18) in place, and the mutation E36G should affect the positioning of all of them, leading to notable structural changes.

Dim56, mutated in amino acid F5, which is part of both the hydrophobic core and the dimeric interface, was predicted to be severely impaired in dimerization and correct folding. Indeed, Dim56 proved to be almost unable to dimerize with a Kd value in millimolar range (430 ± 178 mM). Moreover, the observed CD spectrum indicated major structural changes introduced by the mutation that, in turn, also may have caused the observed dimerization defects.

On the basis of our data with mutations in amino acids predicted to make specific and nonspecific contacts with the DNA we conclude that α-helix III is involved in DNA binding. The observation that both R29Q and R34Q lost the ability to distinguish between A and G in both binding sites (loss of contact mutants) led us to propose that these two amino acids are involved in specific DNA recognition. However, from our experimental data we cannot determine unequivocally which bases in the DNA target are contacted specifically by R29 and R34.

Taken together, our results suggest that CopR binds to the DNA as an HTH protein using α-helix III as the recognition helix.

ACKNOWLEDGMENT

The authors thank O. Kirschstein, IMB Jena, for the opportunity to use the CD spectrometer, E. Birch-Hirschfeld, Institute for Virology Jena, for synthesizing the oligodeoxyribonucleotides, and A. Walter, Institute for Molecular Biology Jena, for the help with the UV-VIS spectrometer to determine the protein concentrations. A.H. thanks the organizers of the FEBS Advanced Course “Frontiers of Protein Structure Prediction 1997” for financing his participation in the course and Tim Hubbard, Alfonso Valencia, Angelika Höninger, and James Cuff for sequence alignments and help with database searches.

REFERENCES


27. Tripos Inc. SYBYL 6.4. 1997. 1699 S. Hanley Road, Suite 303, St. Louis, MO 63144-2913.


