Transcriptional repressor CopR: dissection of stabilizing motifs within the C terminus

Kornelia Kuhn, Katrin Steinmetzer and Sabine Brantl

Author for correspondence: Sabine Brantl. Tel: +49 3641 657576/78. Fax: +49 3641 657520. e-mail: Sabine.Brantl@rz.uni-jena.de

Friedrich-Schiller-Universität Jena, Institut für Molekularbiologie, Winzlerlaer Straße 10, biolitec AG, Winzlerlaer Straße 2a, Jena D-07745, Germany

Abstract: Replication of the streptococcal plasmid pIP501 is regulated by two components, CopR and the antisense RNA, RNAIII. CopR represses transcription of the essential repR mRNA about 10- to 20-fold and, additionally, prevents convergent transcription of sense and antisense RNAs. It has been demonstrated that CopR binds as a preformed dimer. DNA binding and dimerization constants were determined and amino acids were identified that are involved in DNA binding and dimerization. It was demonstrated that the C-terminal 20 aa of CopR are not involved in either activity, but play an important role for CopR stability. Furthermore, it was found that the C terminus of CopR is structured containing a β-strand structure, most probably between the alternating hydrophilic and hydrophobic amino acids 76 and 84 (QVTLELEME). In this study stability motifs within the C terminus of CopR were dissected. Both the cognate and a heterologous (QVTITVTIT) β-strand structure between amino acids 76 and 84 within the C terminus stabilized CopR (CopR derivative CopVT). In contrast, substitution by a predicted α-helix (QVTLKLKMK) or a predicted unstructured sequence (QVTPEPEPE) caused severe and moderate destabilization, respectively. E80 seemed to be the only important C-terminal glutamic acid residue. Deletion of seven C-terminal amino acids from either wild-type CopR or CopVT reduced the half-life to ~ 50% indicating that this C-terminal sequence is a second stability motif.

Keywords: plasmid copy number control, circular dichroism measurements, Gram-positive bacteria, protein stability, structured C terminus

INTRODUCTION

Over the past years, many transcriptional repressors have been studied in detail, e.g. LacI, λ cro, λ repressor, TrpR, TetR, Arc and, in the field of plasmid copy-number control, CopG encoded by pLS1, the structure of which has been resolved recently (Gemis-Ruth et al., 1998). However, in only a few cases have motifs determining the intracellular stability of these control elements been identified and analysed. Pioneering work in this field has been done by the group of R. T. Sauer, who investigated sequences stabilizing the Arc repressor and the unstable N terminus of λ repressor (e.g. Bowie & Sauer, 1989; Parsell & Sauer, 1989; Parsell et al., 1990; Milla et al., 1993, 1994). They found that hydrophilic and polar residues within C-terminal extensions can stabilize otherwise unstable proteins, and even classified amino acids depending on their stabilizing effect. Furthermore, residues closer to the C terminus appeared to play a greater role in determining the proteolytic susceptibility of the proteins than those farther away (Parsell et al., 1990). In all cases, however, the extensions were found to be unstructured and did not interact with the folded portions of the protein (Milla et al., 1993), and the stabilizing function was explained simply with the protection of hydrophobic parts of the proteins by these unstructured hydrophilic/polar residues.

Over the past decade, we became interested in the two components regulating the replication of plasmid pIP501: an antisense RNA which acts via transcriptional attenuation (Brantl et al., 1993; Brantl & Wagner, 1994) and the transcriptional repressor CopR (92 aa, 10-6 kDa). CopR represses transcription of the essential repR mRNA (Brantl, 1994) and prevents convergent transcription of sense and antisense RNAs.
transcription of sense and antisense RNAs (Brantl & Wagner, 1997). Previously, we showed that CopR binds asymmetrically as a preformed dimer at two consecutive major grooves of the DNA (Steinmetzer & Brantl, 1997; Steinmetzer et al., 1998) and identified amino acids involved in DNA binding and dimerization (Steinmetzer et al., 2000a, b). The last 29 aa of CopR are essential for neither DNA binding nor dimerization, but are essential for protein stability (Kuhn et al., 2000). The half-life of CopR amounts to 42 ± 5 min. However, whereas C-terminally truncated CopA20 was, in spite of a drastically shortened half-life, still 100% active in vivo, CopA24 and CopA27 retained only 20% activity. Therefore, we focused our interest on the last 20 aa. Circular dichroism (CD)-difference spectra revealed that this sequence is not unstructured, but part of it, most probably between aa 76 and 84 (QVTLELEME) apparently forms a β-strand with alternating hydrophilic and hydrophobic residues (Kuhn et al., 2000). This β-strand forming sequence is also present in the related repressors CopF (Swinfield et al., 1990) and CopS (Ceglowski & Alonso, 1994), albeit as QVTL...LEME.

Here, we dissect stabilizing motifs within the C terminus of CopR. Two such motifs were found: the amphiphilic β-strand sequence that can be replaced by a heterologous β-strand forming sequence, and the last 7 aa. We demonstrate that neither a region folding predominantly into a z-helix nor a completely unstructured region with the same degree of hydrophilicity can completely compensate for a β-strand.

**METHODS**

**DNA preparation, manipulation and copy-number determination.** Plasmid DNA was isolated from *Bacillus subtilis* as reported previously (Brantl & Behnke, 1992). DNA manipulations like restriction enzyme cleavage and ligation were carried out using the conditions specified by the manufacturer or according to standard protocols (Sambrook et al., 1989). A GeneAmp PCR Kit from Perkin Elmer/Cetus was used for PCRs. DNA sequencing was performed according to the dideoxy chain termination method (Sanger et al., 1977) with a Sequenase kit from Amersham Pharmacia Biotech. Copy numbers of pIP501 derivatives in *B. subtilis* were determined as described (Brantl & Behnke, 1992).

**Construction of *Escherichia coli/B. subtilis* shuttle vectors containing mutated copR genes.** Plasmids pCOPVT, pCOPVTα7 and the plasmids containing individual E to K substitutions were constructed using a single-step PCR on plasmid pCOP7 as template with primer 486-34 (Kuhn et al., 2000) and one of the mutagenic primers described in Table 1. The resulting 500 bp PCR fragment was digested with BamHI and inserted into the single BamHI site of plasmid pPR7 (Brantl & Behnke, 1992).

For the construction of plasmid pCOPEVT, the 500 bp BamHI fragment encoding CopVT was isolated from plasmid pCOPVT and inserted into the single BamHI site of plasmid pPR1-E (Steinmetzer et al., 2000a).

**Construction of *E. coli* vectors for overexpression of mutated copR genes.** A single-step PCR with primer B618-30 (Steinmetzer & Brantl, 1997) and (for pQVT) primer 951-31 (Steinmetzer et al., 2000a) or (for pQPE) primer 8 (Table 1) was used to amplify the mutated copR genes from the *E. coli/B. subtilis* shuttle vectors pCOPVT and pCOPEVT as templates. The amplified fragments were cleaved with BamHI/PstI and inserted into expression vector pQE9 (Qiagen). In this way, the mutated copR genes contain 11 additional 5’ codons encoding Met-Aurg-Sly-Sly-Sly fused to the second codon of copR (Glu).

**Pulse–chase experiments.** These were performed as described previously (Kuhn et al., 2000). In all cases, 100 μl polyclonal antiserum raised against His₆-CopR purified from *E. coli* were used for the precipitation of mutated CopR proteins from the lysates.

**Purification of CopR proteins, determination of protein concentration and electrophoretic mobility-shift assay (EMSA).** Overexpression and purification of the proteins from *E. coli* were performed as described before (Steinmetzer et al., 1998). Protein concentrations were determined by Bradford assays based on calibration curves obtained with the values from acidic hydrolysis of His₆-CopR. EMSA was performed as described previously (Steinmetzer & Brantl, 1997).

**CD measurements.** Purified proteins in 50 mM phosphate buffer (pH 7.5) were dialyzed against 200 mM potassium phosphate buffer (pH 7.5) and 200 μM ethylenediaminetetraacetic acid (EDTA).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Plasmid</th>
<th>Sequence (5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 pCOPVTα7</td>
<td>pCOPPE GAA TCC GGA TCC TTA TTC TGT TAC CGT TAC CGT TAC TGT TAC TGT TCC</td>
<td>AGA TCT GGA TCC TTA TTC TGT TAC CGT TAC CGT TAC TGT TAC TGT TCC</td>
</tr>
<tr>
<td>2 pCOPVT</td>
<td>GAA TCC GGA TCC TCA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
<td>GAA TCC GGA TCC TCA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
</tr>
<tr>
<td>3 pCOPE80K</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
</tr>
<tr>
<td>4 pCOPE82K</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
</tr>
<tr>
<td>5 pCOPE84K</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
</tr>
<tr>
<td>6 pCOPE85K</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
</tr>
<tr>
<td>7 pCOPE86K</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
<td>GAA TCC GGA TCC TTA CAC GAA ATC ATT GCT TTT TTC TCC TGT TAC CGT TAC TGT TCC</td>
</tr>
<tr>
<td>8 pCOPPE</td>
<td>GAA TCC GGA TCC TGC AGT TA C ACG AAA TCA TGG TTTC TTT TCT TCT TCT TCT GGT GCT GCT GGT GGT ACT TGT TCT TGT TCC</td>
<td></td>
</tr>
</tbody>
</table>
buffer containing 150 mM NaCl and 50% (w/w) glycerol were used. The CD spectra were measured at room temperature in the range from 195 to 260 nm with a JASCO model 710 spectropolarimeter at a scan speed of 50 nm min\(^{-1}\) and 1 nm resolution. The path length of the cells used was 0·1 mm. The spectra were recorded as a mean of 10 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.

**RESULTS AND DISCUSSION**

**Design and analysis of C-terminally substituted CopR mutants in vivo and in vitro**

We aimed to answer the following questions by the analysis of half-life and structure of C-terminally mutated CopR proteins: i) does the C terminus need to be acidic to fulfil its stabilizing function? ii) how important are the seven C-terminal amino acids for CopR stability? and iii) is a \(\beta\)-strand forming sequence necessary for CopR stability?

Plasmids pCOPE80K, pCOPE82K, pCOPE84K, pCOPE85K and pCOPE86K were constructed and used for the analysis of single substitutions of acidic glutamic acid residues by basic lysine residues. To investigate the importance of the C-terminal 7 aa residues located downstream of the \(\beta\)-strand forming sequence, pCOPVT\(\Delta\)7 was constructed. To analyse the importance of the previously found \(\beta\)-strand forming sequence QVTLELEME between aa 76 and 84 of CopR, plasmid pCOPVT was constructed encoding a CopR protein where this sequence was replaced by QVTVTVTVT, a sequence which Brack & Caille (1978) had shown to be present in the \(\beta\) conformation in solution. To analyse whether CopVT is able to regulate the copy number of pIP501 in *B. subtilis* as well as wild-type CopR, pCOPEV, a plasmid expressing both regulatory components, RNAIII and the mutated CopVT, was constructed and used for copy-number determinations, since pCOPVT does not express RNAIII. Plasmid pCOPPE was constructed to find out whether the \(\beta\)-strand forming sequence can be replaced by a sequence predicted to form a random coil. In all full-length mutants designed here, we tried to maintain the degree of hydrophilicity within the C terminus as much as possible, since R. T. Sauer’s group demonstrated that hydrophilic C-terminal extensions stabilized proteins more than hydrophobic ones (see Introduction).

All plasmids were transferred into *B. subtilis* strain DB104 and the corresponding strains were used in pulse–chase experiments for half-life determinations (see Table 2). To find out whether CopVT was still able to regulate the replication of pIP501 in *B. subtilis*, copy-number determinations with DB104(pCOPEV) containing a mutant plasmid expressing wild-type antisense RNA, the second regulatory component, and mutated CopVT were performed. Plasmid pCOPEV maintained the wild-type copy-number, indicating that CopVT is functional *in vivo*.

**Table 2. Half-lives and primary C-terminal sequences of wild-type and C-terminally mutated CopR proteins**

<table>
<thead>
<tr>
<th>CopR mutant</th>
<th>Half-life (min)</th>
<th>C-terminal sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CopR</td>
<td>42 ± 5*</td>
<td>IPNEPETEPETEQVTLELEMEEKSNDFV</td>
</tr>
<tr>
<td>Cop(\Delta)7</td>
<td>24 ± 1*</td>
<td>IPNEPETEPETEQVTLELEME</td>
</tr>
<tr>
<td>Cop20-K</td>
<td>1·58 ± 0·42*</td>
<td>IPNEPETEPETKQVTKLKMKKKNSDFV</td>
</tr>
<tr>
<td>CopE86K</td>
<td>50 ± 4</td>
<td></td>
</tr>
<tr>
<td>CopE85K</td>
<td>50 ± 5</td>
<td></td>
</tr>
<tr>
<td>CopE84K</td>
<td>56 ± 4</td>
<td></td>
</tr>
<tr>
<td>CopE82K</td>
<td>40 ± 2</td>
<td></td>
</tr>
<tr>
<td>CopE80K</td>
<td>18 ± 2</td>
<td></td>
</tr>
<tr>
<td>CopVT</td>
<td>111 ± 5 ± 17·5</td>
<td></td>
</tr>
<tr>
<td>CopVTA(\Delta)7</td>
<td>70 ± 5</td>
<td></td>
</tr>
<tr>
<td>CopPE</td>
<td>34 ± 3</td>
<td>IPNEPETEPETEQVTPEPEPEEKSNDFV</td>
</tr>
</tbody>
</table>

* Determined previously (Kuhn *et al.*, 2000).
For the analysis of the structure of mutated C termini, E. coli expression vectors pQVT and pQPE were constructed and used for the purification of His-tagged mutated proteins for CD measurements. Purification of CopVT from an E. coli overexpression strain containing plasmid pQVT was difficult, since the protein tended to aggregate and to bind irreversibly to the HPLC column. Therefore, the yield of purified CopVT was less than 5% of that of the wild-type and other full-length mutated CopR proteins. EMSA with purified CopVT showed wild-type like activity in DNA binding for this mutant (data not shown). All CD spectra revealed the characteristics of proteins containing both α-helices and β-strands, and in the case of CopPE, a substantial amount of random structure as judged by the pronounced minimum at 205 nm. To obtain information about the structural differences of CopPE, CopVT and wild-type CopR in comparison with the previously constructed mutated protein Cop20-K (Kuhn et al., 2000), the difference spectra of CopPE and CopR, CopR and CopVT, and Cop20-K and CopR were calculated (Fig. 1b). The difference spectrum between CopPE and wild-type showed a high amount of random coil, whereas that between Cop20-K and CopR revealed a higher amount of α-helical structures, but also portions of random coil. In contrast, the difference spectrum of CopR and CopVT indicated a moderately decreased α-helical content for CopVT compared with the wild-type, but no significant changes in the β-strand content.

The C terminus of CopR does not need to be acidic

The results obtained previously with Cop20-K (Kuhn et al., 2000; substitution of all Glu residues within the C-terminal 20 aa by Lys residues) did not allow us to decide whether the C terminus of CopR needs to be acidic or not, since the mutant with the basic C terminus also showed an altered C-terminal structure. Therefore, individual substitutions of Glu by Lys residues were analysed. The results of the pulse–chase experiments (determination of half-lives; see Table 2) showed that E85 and E86, located immediately C terminal to the β-strand forming region, could be replaced by K without destabilizing CopR. In contrast, both mutated CopE85K and CopE86K proteins proved to be moderately more stable than the wild-type (50 ± 5 and 56 ± 4 min, respectively). Alteration of single E residues within the β-strand forming region QVTLELEME into K indicated that only E80 seems to play a role in CopR stability, as CopE80K proved to be half as stable as wild-type CopR (18 ± 2 min), whereas CopE84K (56 ± 5 min) was moderately more stable than CopR, and CopE82K (40 ± 2 min) was as stable as the wild-type. Since mutated CopVT protein which does not contain a single E residue within the β-strand forming sequence is nearly threefold more stable than the wild-type (see below), we can conclude that the C terminus of CopR does not need to be acidic. In the wild-type context (QVTLELEME), however, E80 seems to play a decisive role in stabilization.

The C-terminal 7 aa contribute to CopR stabilization by a factor of approximately 2

Deletion of 7 C-terminal amino acids from CopVT (CopVTΔ7) yielded a 1.6-fold reduction of the half-life (70 ± 5 min). A comparison between the half-lives of CopR and CopΔ7 (42 ± 5 and 24 ± 1 min respectively) as well as a comparison between the half-lives of CopVT and CopVTΔ7 (111 ± 5 and 70 ± 5 min respectively) shows that the 7 C-terminal amino acids (EKSNDVF) contribute to CopR stability by a factor of 1.6-1.8. Previously, we had also shown that mutated protein CopA5, with a half-life of 31 ± 6 min, is moderately less stable than wild-type CopR. Four of the C-terminal 7 aa and two of the C-terminal 5 aa are hydrophilic, which does support the role of hydrophilic amino acids in protecting more hydrophobic regions of the protein against proteolytic attack, as suggested by the publications of R. T. Sauer’s group (Milla et al., 1993; Parsell & Sauer, 1989; Parsell et al., 1990).
An amphiphilic β-strand structure within the C terminus stabilizes CopR, whereas an α-helix or an unstructured region does not

CD measurements with purified CopVT suggested that this protein with the sequence QVTTVTVT between aa 76 and 84 which has a higher propensity to form an amphiphilic β-strand has a wild-type-like structure (Fig. 1a). Pulse–chase experiments showed that it is nearly threefold more stable in vivo than the wild-type protein [Table 2, half-lives of CopR (42 ± 5 min) and CopVT (111.5 ± 17.5 min)]. As shown before, replacement of this sequence by a sequence that folds predominantly into an α-helix (Cop20-K; QVTLKLKMK) significantly destabilized CopR (half-life of 1.6 ± 0.4 min; Kuhn et al., 2000). At that time, however, it was not clear whether this effect was mainly due to structural reasons or to the exchange of an acidic for a basic C terminus. To prove whether an unstructured region can compensate for the β-strand forming region, we analysed mutated CopPE protein containing the sequence QVT-PEPEP between aa 76 and 84. CD measurements suggested that this mutant contains a random coil region (Fig. 1) and pulse–chase experiments showed that it was less stable than the wild-type (34 ± 3 min compared with 42 ± 5 min), although not to the extent exerted by Cop20-K. The relatively moderate destabilization of CopPE can be explained by the presence of both the stabilizing C-terminal 7 aa and the decisive E80 within the β-strand forming sequence, each of which contributes by a factor of ~2 to CopR stability (see above). Furthermore, it cannot be excluded completely that one or two of the proline residues can form hydrophobic contacts which are made by leucine or valine residues in the case of wild-type CopR or CopVT, respectively, since for example in 434 repressor N terminus, proline is part of the hydrophobic dimeric interface (Aggarwal et al., 1988). In contrast, the significant destabilization of Cop20-K seems to be predominantly due to structural alterations (β-strand into α-helix and random coil): the stabilizing 7 C-terminal aa are present in this mutant and replacement of E85 or E86 by K as present in Cop20-K was found to have a stabilizing rather than a destabilizing effect.

These results indicate that the amphiphilic β-strand is an important stabilizing motif within the C terminus of CopR and that its substitution by an α-helix or an unstructured region with the same degree of amphiphilicity cannot, or can only partly, compensate for its stabilization function.

We suggest that an interaction of the β-strand forming sequence with amino acids located at the N terminus and/or in the central portion of the protein leads to the stabilization of CopR. Among these contacts, at least one ionic contact between E80 and a basic residue should be expected, whereas the other contacts have to be hydrophobic ones involving one or more of the hydrophobic aa within the β-strand (V77, L79, L81 or M83). Experiments are under way to prove this hypothesis.

ACKNOWLEDGEMENTS

We acknowledge E. Birch-Hirschfeld (Institute for Virology, Jena) for synthesizing various oligodeoxynucleotides and Waltraud Scheiding for the help in protein purification. This work was supported by grant BR 1552/4-2 from the Deutsche Forschungsgemeinschaft (to S.B.).

REFERENCES


with chain-terminating inhibitors. Proc Natl Acad Sci USA 74, 5463–5467.


