Structural basis of bifunctionality of *Sinorhizobium meliloti* Clr, a cAMP and cGMP receptor protein

Laura Werel[a§], Neda Farmani[b,c§], Elizaveta Krol[b,c], Javier Serrania[b,c], Lars-Oliver Essen[a*] and Anke Becker[b,c*]

[a]Department of Chemistry, Philipps-Universität Marburg, Marburg, Germany

[b]Department of Biology, Philipps-Universität Marburg, Marburg, Germany

[c]Center for Synthetic Microbiology (SYNMIKRO), Philipps-Universität Marburg, Marburg, Germany

[§]Authors contributed equally to this work

*Corresponding authors: Lars-Oliver Essen (essen@chemie.uni-marburg.de) and Anke Becker (anke.becker@synmikro.uni-marburg.de)

Keywords: nucleotide second messenger, CRP-like protein family, allosteric control, Clr regulon, root nodule symbiosis.

ABSTRACT

In bacteria, Crp-Fnr superfamily transcription factors are the most ubiquitous receptor proteins of 3',5'-cyclic adenosine monophosphate (cAMP) and 3',5'-cyclic guanosine monophosphate (cGMP). The prototypic *Escherichia coli* CAP protein represents the main CRP subclass and is known to bind cAMP and cGMP, but to mediate transcription activation only in its cAMP-bound state. In contrast, both cyclic nucleotides mediate transcription activation by CRP subclass G protein Clr of *Sinorhizobium meliloti*. We present crystal structures of apo-Clr, and Clr•cAMP and Clr•cGMP bound to the core motif of the palindromic Clr DNA binding site (CBS). We show that both cyclic nucleotides shift ternary Clr•cNMP•CBS-DNA complexes to almost identical active conformations. Unlike the situation known for the *E. coli* CAP•cNMP complex, in the Clr•cNMP complex, the nucleobases of cGMP and cAMP are in the *syn*- and *anti*-conformation, respectively, allowing a shift to the active conformations in both cases. Isothermal titration calorimetry measured similar affinities of cAMP and cGMP binding to Clr in presence of CBS core motif DNA (K_D^{cNMP} 16 µM). However, different affinities were determined in absence of this DNA (K_D^{cGMP} 24 µM; K_D^{cAMP} 6 µM). Sequencing of Clr co-immunoprecipitated DNA as well as Electrophoretic Mobility Shift and promoter-probe assays expanded the list of experimentally proven Clr-regulated promoters and CBS. This comprehensive set of CBS features conserved nucleobases, which are in agreement with the sequence readout through interactions of Clr amino acid residues with these nucleobases, as revealed by the Clr•cNMP•CBS-DNA crystal structures.

IMPORTANCE

Cyclic 3',5'-adenosine monophosphate (cAMP) and cyclic 3',5'-guanosine monophosphate (cGMP) are both long known as important nucleotide second messenger in eukaryotes. This is also the case for cAMP in prokaryotes, whereas a signaling role for cGMP in this domain of life has been recognized only recently. Catabolite repressor proteins (CRPs) are the most ubiquitous bacterial cAMP receptor proteins. *Escherichia coli* CAP, the prototypic transcription regulator of the main CRP subclass, binds both cyclic mononucleotides, but only the CAP•cAMP complex promotes transcription activation. In contrast, CRP subclass G proteins studied so far, are activated by cGMP,

or both by cAMP and cGMP. Here, we report a structural analysis of the bifunctional cAMP- and cGMP-activatable Clr from *Sinorhizobium meliloti*, how binding of cAMP and cGMP shifts Clr to its active conformation, and the structural basis of its DNA binding site specificity.

INTRODUCTION

3',5'-cyclic adenosine monophosphate (cAMP) is one of the most ubiquitous nucleotide second messengers in prokaryotes and eukaryotes. It was the first second messenger to be described, initially for its role in hormone-dependent signal transduction by eukaryotes (1, 2), later for catabolite repression in bacteria like *Escherichia coli* (3). Since then, a myriad of biological processes regulated by cAMP have been reported in bacteria, such as carbon metabolism, biofilm formation, type III secretion, virulence, and symbiosis (4, 5). Another cyclic mononucleotide second messenger widespread in eukaryotes is 3',5'-cyclic guanosine monophosphate (cGMP) (6). Involvement of cGMP in bacterial regulation has been recognized only recently (7).

In bacteria, cyclic AMP receptor proteins (CRPs) are among the best characterized transcription factors and model for allosteric regulation (3, 8-10). Physiological effects of gene regulation by CRP-like proteins were characterized in various bacterial species. They are highly versatile, controlling expression of at least over 200 genes in *E. coli* and almost as many in *Mycobacterium tuberculosis* and *Corynebacterium glutamicum* (11-*15*).

CRPs belong to the superfamily of Crp-Fnr transcription regulators, which are composed of an N-terminal nucleotide binding domain and a C-terminal helix-turn-helix motif (16). These domains are each conserved in eukaryotes and prokaryotes (4, 17). Upon DNA binding, the C-terminal domain of CRPs is able to interact with the bacterial RNA polymerase (RNAP) and to modulate promoter activation (18-20). Most of the CRPs act as transcription activators, few as repressors (6).

CRP-like proteins of the Crp-Fnr superfamily were further divided into the main CRP, and CRP F and G subclasses (16). So far, all structurally characterized CRPs belong to the main CRP subclass, including *E. coli* CAP (21). These proteins are activated by cAMP only (16). Only two members of the CRP G subclass have been functionally studied. CrgA from *Rhodospirillum centenum* and Clr from *Sinorhizobium meliloti.* CrgA appeared to be solely activated by cGMP with

very little cAMP-binding activity, whereas Clr is distinguished from any of the other known CRP homologues by its ability to be activated by both cAMP and cGMP (22, 23). cGMP binding has been described for *E. coli* CAP as well, albeit without stimulating DNA binding (24).

The soil-dwelling α-proteobacterium *S. meliloti* is capable of fixing atmospheric nitrogen in a symbiotic relationship with leguminous plants of the genera *Medicago*, *Melilotus* and *Trigonella*. The symbiotic program involves root hair infections and formation of root nodules, harbouring the nitrogen-fixing bacteria (25, 26). The *S. meliloti* Rm1021 and Rm2011 genomes encode 13 CRP-Fnr-like paralogs, including the cAMP-independent FixK regulators of microoxic respiration and nitrogen-fixation genes (3), and cAMP- and cGMP-controlled Clr (27). Eight of those are members of the CRP superfamily. Bifunctionality of Clr is especially interesting given that these *S. meliloti* genomes contain an exceptionally high number of 28 putative adenylate/guanylate cyclases (AC/GCs) genes (28). Clr and the AC/GCs, CyaD1, CyaD2 and CyaK, are implicated in repression of secondary infections of *Medicago sativa* roots (5). Moreover, Clr overexpression resulted in reduced swimming motility and increased succinoglycan production, which are traits relevant at early stages of the root nodule symbiosis (23).

In our study, we present crystal structures of apo-Clr and Clr•cNMP•DNA complexes, which to our knowledge is the first report of a subclass G CRP structural analysis. Our analysis unravels the molecular basis of Clr bifunctionality. Furthermore, we expand the set of experimentally proven Clr-regulated promoters and CBS. These feature conserved nucleobases which are in accordance with the sequence readout revealed by the Clr•cNMP•CBS-DNA crystal structures.

RESULTS

ChIP-seq assisted screening for candidate Clr binding sites. To identify Clr-cAMP and Clr-cGMP binding sites genome-wide in the DNA of *S. meliloti* Rm2011, we used chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq). For this purpose, we first generated plasmid pWBT-Clr-CF and introduced it into *S. meliloti* strain Rm2011 Δ*clr*. pWBT-Clr-CF carries IPTG-inducible *clr-3xflag*, encoding Clr C-terminally fused to a 3xFLAG-tag. *In vivo* functionality of Clr-3xFLAG was confirmed by its ability to activate plasmid-borne *egfp* reporter

gene fusions to promoters of the previously identified target genes *SMc02178* and *SMb20906* (23) in Rm2011 Δ*clr* in presence of externally added cAMP (Fig. S1). For α-FLAG antibody mediated chromatin immunoprecipitation, Rm2011 Δ*clr* carrying pWBT-Clr-CF was cultivated in TY complex medium or MOPS-buffered minimal medium (MM), and supplemented with either cAMP or cGMP. Deep sequencing of DNA enriched for Clr-bound regions and control samples representing sheared DNA isolated from cell lysates prior to immunoprecipitation yielded 3.3 to 3.9 million reads per sample (Table S1A). ChIP-seq data was processed using CLC-Genomics workbench to determine peak locations and peak shape score values, which reflect enrichment of the respective regions (Table S1C-F). Consolidation of our data obtained for the four samples yielded 873 distinct ChIP-enriched genomic regions, further referred to as ChIP peaks (Table S1G). Consistent with our previous observation that both cAMP and cGMP promote Clr binding to DNA (23), 40 out of 50 top-score ChIP-enriched regions were detected in samples from cultures supplemented with either of the two cNMPs (Table S1G). Moreover, peak shape score values correlated well between cAMP and cGMP samples (correlation factors 0.92 and 0.91 in TY and MM media, respectively) (Fig.1A).

Out of 417 ChIP peaks identified within intergenic regions, 160 were located between two divergently transcribed genes, 229 between two genes transcribed in the same direction and 28 between two convergently transcribed genes (Fig.1B). The remaining ChIP peaks were situated within rRNA or tRNA genes and protein coding regions (Fig. 1B). To search for Clr binding sites (CBS) in ChIP-enriched regions, we extracted 400 bp DNA sequences surrounding the peak centers. Using the FIMO online tool (29), these regions were scanned for the core Clr binding site motif GTNNCNNNNGNNAC. This core motif was derived from the consensus sequence HGTYHCNNNNGRWACA, which represents six experimentally verified CBSs (23). Sequences matching the core motif were found in 102 ChIP-enriched regions, corresponding to a total of 107 putative CBSs (Fig. 1C and Table S1H), out of which 87 were located in intergenic regions (Table S1H). A FIMO search with relaxed settings (p-value 0.001) identified additional 693 putative CBSs with one mismatch in the core motif (relaxed-consensus CBSs) in 431 ChIP-enriched regions

(Table S1I). Out of these, 89 putative CBSs were found in 51 ChIP-enriched regions, which already contained a strict consensus sequence.

As expected, all previously characterized CBS sequences in the promoter regions of *SMc02178*, *SMc00653*, *SMc01136*, *SMc04190*, *cyaF2* and *SMb20906* (5, 23) were found within the ChIP-enriched regions. Among genes adjacent to ChIP peak-containing intergenic regions, 35 genes were found (Table S1J) that were previously identified in transcriptome experiments to be upregulated upon overexpression of *clr* or adenylate cyclase gene *cyaJ* (23, 30). 25 of these genes contained putative consensus and/or relaxed-consensus CBSs in their upstream non-coding regions. The remaining 10 genes lacked putative CBSs in their vicinity. Remarkably, these 10 genes were only detected as being transcriptionally activated upon *cyaJ*, but not *clr* overexpression.

Clr-cAMP and Clr-cGMP specifically bind DNA target sequences *in vitro.* To characterize putative CBS sequences identified within ChIP-seq-enriched DNA regions and located in upstream non-coding regions, we performed electrophoretic mobility shift assays (EMSAs) for 62 genomic regions (Fig. 2A and Fig. S2). The corresponding DNA fragments that produced a complete electrophoretic mobility shift in presence of Clr and cAMP were defined as regions with a strong CBS, whereas DNA fragments that were only partially shifted were defined as containing a weak CBS (Fig. 2B). Of the 62 CBS candidates tested, 24 were classified as strong and 24 as weak CBSs, while the remaining failed to mediate a detectable band shift in the EMSAs (Fig. 2A and Fig. S2). Furthermore, Clr-cAMP produced stronger band shifts than Clr-cGMP. Not all DNA regions containing putative CBS matching the consensus produced strong band shifts, which indicates that non-conserved nucleotides within the CBS and its sequence environment might contribute to Clr-cNMP binding. For strong binding sites, we noted that the left and right halves of the palindrome tended to be enriched for pyrimidine (C/T) and purine base nucleotides (G/A), respectively (Table S1K). Conversely, putative CBSs mediating a weak or no band shift lacked a clear nucleotide preference at the non-conserved positions. Almost all DNA fragments causing a strong band shift included a CBS sequence that matches the core motif. A notable exception is the DNA fragment covering the *SMc05008* promoter region, as this region contains three overlapping putative CBSs,

each with one mismatch in the core motif (Fig. 3).

To verify the CBS location within DNA regions that showed strong Clr-cNMP binding, we exchanged the conserved GT in the left half of the palindrome to CA. This modification dramatically reduced Clr-cNMP binding, thus confirming the predicted CBS within these regions (Fig. S2). Out of the three overlapping putative CBS in the *SMc05008* promoter region, only mutations in conserved nucleotide positions of the overlapping putative second and third CBS inhibited Clr-cNMP binding (Fig. 3). Band shift assays with synthetic DNA fragments containing either one of the putative CBSs identified in the *SMc05008* promoter region, suggested that only the second putative CBS conferred Clr-cNMP binding (Fig. S2D). This indicates that a C→T transition at position 6 of the CBS core motif is not detrimental *per se* for Clr-cNMP binding to the DNA.

Target promoters are regulated both by Clr-cAMP and Clr-cGMP *in vivo.* All tested CBS candidates were further investigated for Clr-cNMP dependent promoter activation *in vivo*. In case the CBS was located in non-coding regions between two opposingly transcribed genes, both putative promoter regions were tested. Plasmid-borne copies of promoter regions followed by up to 75 bp of the downstream coding regions or non-coding RNA genes fused to *egfp* were introduced to *S. meliloti* or *E. coli* for promoter activity assays.

Reporter construct-mediated fluorescence was determined in *S. meliloti* Rm2011 wild type and Rm2011 Δ*clr* upon addition of either cAMP or cGMP (Fig. 4A, Fig. S2E and Table S2). This analysis confirmed Clr-cNMP-dependent activation of 20 promoters by at least 1.5-fold either with cAMP or cGMP added. These included only promoter regions that showed a strong band shift in the EMSAs. As expected, GT→CA mutation of the first two nucleotides of the core CBS motif abolished or substantially decreased Clr-cNMP dependent promoter activation (Fig. S2E and Table S2). Regarding the three overlapping CBS in the *SMc05008* promoter region, only nucleotide exchanges that diminished Clr-cNMP binding in the EMSA also abolished promoter activation by Clr-cNMP (Fig. 3). For the majority of CBS-containing promoter fragments tested, induction by cGMP was up to four times higher than by cAMP. In total, we identified 11 novel Clr-cNMP-activated genes in addition to nine Clr-cNMP-activated genes reported previously (23, 30). The majority of

the 20 confirmed Clr-cNMP regulated genes encode hypothetical proteins and, notably, eight of the 20 genes encode small proteins of less than 100 amino acids. Ten Clr-cNMP-activated genes encode proteins with an N-terminal signal peptide. Moreover, Clr-cNMP was found to directly induce transcription of *SMelC181* encoding a small non-coding RNA of unknown function. This RNA was previously identified in an RNA-seq study (31).

Furthermore, Clr-cNMP-mediated transcription activation of overall 17 promoter regions with confirmed and putative Clr-binding motifs was analyzed in *E. coli* as a heterologous host (Fig. 4B and Table S3). We used *E. coli* strain BTH101, which is unable to produce cAMP. This strain was equipped with plasmid-borne *clr*, co-expressed with either the *S. meliloti* adenylate cyclase gene *cyaG1* or the *Synechocystis* sp. guanylate cyclase gene *cya2* under the control of an IPTG-inducible promoter. Promoter-reporter constructs were introduced on a second plasmid. We determined previously that in *E. coli* BTH101, cAMP/cGMP production alone was insufficient to activate the *S. meliloti* Clr-cNMP target promoters, and therefore *E. coli* CRP did not substitute for Clr (23). Out of the 17 promoters tested, eight were activated by Clr-cNMP in *E. coli* by at least 1.5-fold either by cAMP or cGMP production. Including the three previously described candidates *SMb20906*, *SMc00653* and *SMc04190* (23), a total of 11 promoters were activated by Clr-cNMP in *E. coli*. Promoters that were not clearly activated by Clr-cNMP in *S. meliloti* Rm2011 also showed no activation in *E. coli* BTH101 (Tables S2 and S3). Contrarily, five promoters (*SMc00854*, *SMa6349*, *SMc06520*, *SMc02177* und *SMc00925*) that were activated in *S. meliloti* Rm2011 were not activated in *E. coli* BTH101, suggesting that these promoters might require additional *S. meliloti*-specific factors for Clr-cNMP-mediated activation.

To evaluate promoter activation by Clr dependent on endogenously produced cNMPs, we compared the promoter-probe constructs conferring EGFP fluorescence between Rm2011 and its derivatives lacking either all AC/GC genes (cya^0) or *clr* (Fig. 4C). To generate the cya^0 strain, all 28 annotated class III AC/GC-encoding genes (28) were sequentially deleted from the Rm2011 genome. The strains were grown in medium without added cNMPs. Under these conditions, eight and seven of the 20 promoter-probe constructs mediated at least 1.5-fold ($p < 0.05$) reduced

fluorescence levels in the Δ*clr* and *cya*^{0} strains, respectively, compared to the wild type (Fig. 4C and Table S4). Lack of the *clr* gene resulted in a much stronger reduction in the promoter activities of *SMc00854*, *SMc02178* and the sRNA gene *SmelC181* than absence of the 28 AC/GC genes.

To estimate the effect of CBS location relative to the transcription start site (TSS) on promoter activation, we gathered *S. meliloti* TSS positions, determined previously in Rm2011 and Rm1021 wild type strains (32, 33) and Rm1021 derivative RFF625c lacking all *ecf* sigma factors, and determined the CBS-TSS distances (Table S1H, I). For most promoters activated by Clr-cNMPs, CBS-TSS distances of -42 to -76 were found (Fig. 5). These promoters are thus comparable to CAP-dependent class I and class II promoters in *E. coli* (34-36). In *E. coli*, these promoters are subdivided into three classes: promoters with a CAP binding site roughly 60 to 90 bp upstream of the TSS (class I), close to the -35 element (class II), and at position -91 or further upstream (class III); the latter requiring additional binding of another transcription regulator closer to the promoter region for activation (34-36).

Clr affinity for cGMP increases in presence of DNA. As EMSA and promoter activation studies suggested cNMP-inducible binding of CBS-DNA by Clr, we further analysed the binding behaviour of Clr using isothermal titration calorimetry (ITC, Fig. S3). A synthetic 33 base pair DNA oligomer was designed based on the core CBS motif as binding target for Clr. It contains a double strand break and was subsequently used in crystallisation experiments (see Fig. 7). The obtained thermograms are consistent with a simple model assuming one binding site per Clr molecule, i.e., no cooperativity between the binding sites. A similar behaviour has been described for *E. coli* CAP (18**,** 37**,** 38) and *M. tuberculosis* CRP (38). The 'active' form of CAP is represented by a CRP dimer·$cAMP_2$ complex (37**,** 39). The secondary binding site proposed for *E. coli* CAP and *M. tuberculosis* CRP located in the *C*-terminal HTH DNA binding domain is not conserved in Clr and no binding could be observed in ITC, HDX-MS and crystallographic experiments (40). As the affinity for the secondary binding site is considerably above physiological concentrations its biological relevance has recently been dismissed (9**,** 39**,** 41**,** 42).

The affinity of cAMP binding by Clr is mostly not affected by the presence of cognate CBS-DNA

with K_D values of 6-7 µM (Table 1). Formation of the ternary Clr•cAMP•CBS-DNA complex is enthalpically driven with an opposed loss of entropy, whereas that of the binary Clr•cAMP complex relies both on enthalpic and entropic contributions. In contrast, cGMP binding by Clr alone proceeds endothermically and with apparently lower affinity (K_D ~24 µM). Compared to the Clr•cAMP•CBS-DNA complex, ternary complex formation with cGMP gives similar affinities, and also relies on both contributions. For Clr alone, no DNA binding could be detected under ITC conditions. In the apo state of *E. coli* CAP, the cAMP and CBS-DNA binding domains appear as quasi-independent dynamic units, while the intra- and inter-subunit correlations increased in the holo state (43). Our ITC data indicate a profound difference only for the binary Clr•cNMP complexes, which may affect formation of ensembles capable of binding to CBS-DNA. According to size exclusion chromatography, apo-Clr forms a monomer-dimer equilibrium in solution that favours the monomeric state with 86%. For comparison, the best studied Clr homolog is *E. coli* CAP, which is able to bind both cAMP and cGMP, but is only activated by cAMP for DNA binding (44). The lack of activation in response to cGMP has also been found for *M. tuberculosis* Rv3676 and has been assumed to be a generic trait of Crp-like proteins (41).

Both cAMP and cGMP are able to shift Clr to its active conformation. Unlike apo-Clr and its binary complexes, we found that ternary Clr•cNMP•CBS-DNA complexes crystallize within a few hours in the same, orthorhombic crystal form. Given their intrinsic high anisotropic diffraction characteristics we employed extensive optimisation screening to yield crystals diffracting to 2.8 and 3.1 Å for the cAMP- and cGMP-bound Clr•CBS-DNA complex, respectively (PDB accession code 7PZA and 7PZB, Table S5). The structures show a Clr dimer that is bound to 26 bp CBS-DNA containing a central double strand break. The individual subunits fold into two domains linked by helix α5 (residues P123-L149), which is the Clr dimerization interface. The α5 helices of the dimer form a coiled-coil mainly driven by hydrophobic interactions, with T140 and T141 as notable exceptions. Residues M1-Q121 belong to the cNMP binding domain, the larger of the two. It is formed by four α helices and seven β strands in an antiparallel orientation of a β-sandwich encasing the cNMP binding site. A helix-turn-helix (HTH) DNA binding domain consisting of four α helices

and two β strands is located at the C-terminus (Y150-D234). The cNMP binding domain is universally conserved throughout cNMP receptor proteins, as well as protein kinases and ion channels, with sequence identities above 20% even for mammal representatives (17). The HTH motif of the DNA binding domain is prevalent in 116 structures so far deposited in the PDB according to InterPro classifications (45**,** 46). Thus, both domains perform their function as core components of a variety of proteins adapted to fulfil different biological tasks. Accordingly, the overall fold of Clr is highly similar to its *E. coli* homologue, the catabolite activator protein CAP (1CGP) (47) and even more so to the *M. tuberculosis* (3MZH) and *C. glutamicum* CRPs (4CYD) (48**,** 49). The latter two served as templates for molecular replacement.

The CBS-DNA interacts with its symmetry mate in the crystal lattice and is therefore well resolved. When comparing the ternary complexes, the Clr•cNMP structures as well as the CBS-DNA show a high conformational congruity with pairwise r.m.s.d. values below 0.6 Å. This further supports our notion that both cyclic nucleotides shift ternary Clr•cNMP•CBS-DNA complexes to almost identical active conformations. The cNMP binding site located in the N-terminal domain is stoichiometrically occupied by cAMP and cGMP, respectively (Fig. 6A and B, and Fig. S4). The binding mode of the phosphate and sugar moieties is identical, involving G85 and E86 within α4 as well as R95 and S96 of the following loop. The sidechain of S96 is flipped towards the phosphate only in the cGMP-bound complex. The nucleobase of cGMP is flipped to the *syn*-conformation, while it occupies the *anti*-conformation in the Clr•cAMP complex. Both nucleobase conformations allow for interaction with helix α5 residue T140 of the same Clr subunit and T141 of the adjacent Clr subunit. While cAMP interacts via its primary N^6-amine with both T140 and T141, cGMP forms hydrogen bonds between T140 and its C6 carbonyl group as well as between T141 and the N^2-amine. The latter interaction between the N^2-amine of cGMP and T141 of the adjacent molecule is a substantial difference between Clr and the inactive cGMP complex of *E. coli* CAP (50). Interaction of the cAMP nucleotide with both α5 helices has been described to trigger a coil-to-helix transition in the *C*-terminal part of the hinge and shift the HTH domain towards its active conformation for CAP (51). So far, cGMP appeared to be incapable of such an interaction with the second helix for

283 promoting CRP activation (44). In case of Clr both the interaction and arrangement of the full α5
284 helix is resolved by the structure, which is in an agreement with Clr-DNA binding *in vitro* and
285 activation of promoter activity *in vivo* by both cNMPs. Moreover, mapping the relative hydrogen-
286 deuterium exchange between apo and holo states by HDX-MS experiments shows that upon
287 nucleotide addition the *C*-terminal half of the helix gets shielded from exchange. This further
288 supports the notion, that both nucleotides are capable to trigger the coil-to-helix transition in the
289 linker to the C-terminal helix-turn-helix domain (Fig. 6C and D). This shielding is not only due to
290 nucleotide interactions with T140 and T141, as even fragments, containing the last five residues of
291 the α5 helix, show a clear decline in solvent accessibility for nucleotide-bound states. Helix α6 is
292 shortened in response to the hinge rearrangement and interacts with the former in an angle of
293 about 40° in the active conformation (11). This interaction is stabilised in Clr by a formerly unknown
294 chloride-binding site. The relative hydrogen-deuterium exchange comparing the apo and holo
295 states also shows a slightly stronger decline in deuterium uptake within the binding pocket for the
296 cAMP-bound compared to the cGMP-bound state. Some uptake differences within helix α7 and α8
297 of the HTH domain can be observed (Fig. S4). Upon nucleotide binding the deuterium exchange
298 increases in this region, indicating some degree of conformational change. This is in agreement
299 with structures published for *E. coli* CAP, where the active conformation is achieved by a 60°
300 rotation of α8 (44, 52, 53).

301 **The crystal structures of nucleotide bound Clr shed light on its DNA specificity.** Clr binds
302 to a target DNA of 33 bp length via electrostatic interactions and hydrogen-bonding to the major
303 groove. The surface of the HTH DNA binding domain has a strongly positively charged patch with
304 which it interacts with the DNA phosphate backbone (Fig. 7A). Clr has a likewise positively charged
305 cavity in each subunit for accommodating cNMP. Helix α8 is inserted in the major groove and
306 interacts with hydrogen bonds over six bases as depicted in Fig. 7B and C. The binding results in
307 a 40° and 32° kinking of the DNA helix at the longer and shorter DNA oligomers, respectively, which
308 is eased by the engineered double strand break at the twofold symmetry centre. The entropic cost
309 associated with DNA bending is overcompensated by establishing further interactions. As a result,

the flexibility of the flanking regions of the recognition sequence might be distinguished by Clr, adding another layer of specificity (54, 55).

The electron density of the duplex DNA that was annealed from four oligonucleotides (Fig. 7D) defines the 14 bp of the 5'-oligonucleotides of the core as well as 19 (17) bp of the two 3'-oligonucleotides. The base readout occurs at most conserved positions of the core CBS motif (Fig. 7D) (23). Q184 of α7 and R195 and N199 of α8 are interacting with the ^{9}TGT triplet of the 14-mers, whereas K197 and S194 of the same helix, as well as L152 within the loop *C*-terminally of α5 are hydrogen-bridged to the ^{4}GGT triplet of the 19-mer. For *E. coli* CAP mainly R180, E181 and R185 have been identified to interact with the recognition sequence (56, 57). In Clr, R180 corresponds to R195, E181 is not conserved and R185 is conserved as R200, but fails to interact with the DNA as it adopts a different rotamer. Accordingly, CRP-like proteins in the same phylogenetic subclass as *E. coli* CAP have a highly conserved RE(T/M)VGR motif in their α8 helix, whereas CRP subclass G, to which Clr belongs, is significantly more diverse in this region (16).

CRP proteins interact with the bacterial RNAP via the stacked β-strands within the HTH domain that has been termed AR1 in class I promoters (56, 58) or the top of the beta barrel within the cNMP binding domain, which has been labelled AR2, in class II promoters (59, 60). For Clr AR1 three of the four beta strands are present, whereas the N-terminal loop of the motif is turned more towards the protein core (Fig. S4). Crystal structures of CAP in complex with DNA and the αCTD domain of the RNAP show very little flexibility within the interface and surprisingly only one hydrogen bond between T158 of CAP and T285 of the αCTD (PDB accessions 5CIZ, 3N4M and 1LB2). The interface is largely hydrophobic, a feature that is retained in Clr. The activating region AR1 has a significantly divergent orientation in its *N*-terminal part for Clr. This does prevent interaction with the *E. coli* RNAP *in vivo,* which we infer from promoter activation by Clr in *E. coli* (see above). The overall conservation of the 22 residues comprising AR1 is rather low with three identical and three similar amino acids. In class II promoters the RNAP interacts with the CRP on two additional sites: AR1 of the second dimer symmetry unit and AR2 of this unit (3). The interactions with AR2 involve the beta and omega subunit as shown in crystal and cryo-EM structures from *Thermus*

337 *thermophilus* and *E. coli* (60, 61). The interactions of Clr within the omega unit are conserved as
338 G47 and D48 (G14 and E15 in *T. thermophilus*), whereas the area involved in the interaction with
339 the beta subunit in *T. thermophilus* (mainly through E8) deviates in a way that allows no deductions
340 on whether it would interact in the same way.

341 ## DISCUSSION

342 Our study was focused on the bifunctionality of Clr. We confirmed that Clr is able to bind cAMP
343 and cGMP *in vitro* and characterized the binding properties of this transcription regulator for both
344 these cyclic nucleotides and for its DNA binding site. In our ITC measurements cAMP bound to Clr
345 with about four times higher affinity than cGMP (K_D^{cGMP} = 24µM; K_D^{cAMP} = 6 µM). For CAP from *E.*
346 *coli*, values in the same range have been reported, the affinity for cAMP (10 µM) also being greater
347 than that of cGMP with 16 µM (9, 40). However, we showed that the difference between the
348 nucleotide affinities of Clr diminished as soon as cognate DNA was present, showing that ternary
349 complex formation by both effector molecules is similarly driven.

350 The regulation of cAMP receptor proteins depends on an equilibrium between an active form
351 that is capable of DNA interaction and an inactive form that is not. The equilibrium is shifted towards
352 the active conformation by effector binding (18). For Clr, both our crystal structures and the HDX
353 experiments clearly prove that Clr undergoes this switch independently of the purine base of cNMP.
354 As the cNMP binding site does not deviate significantly from that of the *E. coli* homologue, Clr might
355 have a lower energy barrier for the shift in population towards the active conformation. This is
356 further supported as similar equilibrium shifts have been introduced via mutations of *E. coli* CAP
357 before (62). Cyclic AMP binding initiates the population shift by interacting with residues of the
358 central helices of both subunits (CAP:T127 and CAP:S128) (63). This in turn leads to a coil-to-helix
359 transition in the *C*-terminal part of the helix, resulting in a DNA binding domain in its active
360 conformation (11, 52, 64, 65). Contrary to any previous findings in *E.* coli, we clearly established
361 that Clr performs an identical binding movement via a ^{140}TT motif for both cNMPs. As cGMP is
362 bound in a *syn* conformation the interactions can be retained. We were able to show the resulting
363 coil-to-helix switch of the α5 helix by HDX-MS in solution due to a decline in deuterium uptake,

364 further proving the bifunctional nature of Clr. Clr residues G85, E86, R95 and S96, which are also
365 part of the binding pocket, are largely conserved in CAP as well and responsible for the high-affinity
366 nucleotide binding (66).

367 The C-terminal helix-turn-helix motif of CRPs recognizes its cognate consensus DNA motif via
368 major groove readout (43). Binding of the Clr dimer to two symmetrically related motifs results in a
369 kink of the DNA by 40° and 32°. For CAP, DNA kinks of even 110° have been reported and modified
370 promoters containing an inherent kink were shown to function CRP-independent, which highlights
371 the importance of this particular feature (67, 47, 57, 68, 69). Our isothermal titration calorimetry
372 measurements of DNA binding gave dissociation constants of 6.0 and 4.5 µM and free energy
373 changes of -7.1 and -7.3 kcal/mol for cAMP and cGMP, respectively. This makes Clr a transcription
374 factor on the weaker side of the binding spectrum. However, stronger does not always equal better,
375 as the binding needs to be reversible as well (54). For duplex-DNA of similar length CAP
376 dimer·$cAMP_2$ binding affinities of about 10 µM have been reported, which suggests similar binding
377 strengths for CAP and Clr (69).

378 Clr interacts with the nucleobases of the DNA core motif via hydrogen bonds. The residues
379 involved in sequence readout are exactly contacting the bases within the CBSs as derived from
380 genome-wide ChIP-seq and EMSA data of this work. The bifunctionality of Clr was demonstrated
381 *in vivo* as well. Virtually all promoters directly targeted by Clr were activated by Clr-cAMP and Clr-
382 cGMP in both *S. meliloti* and *E. coli*. The arrangement of three overlapping imperfect CBS core
383 motifs in the *SMc05008* upstream region is intriguing. It suggests a binding mode different from
384 binding of a single CBS or that differential binding of the individual sites might generate a different
385 regulatory output. In *S. meliloti*, we found that externally added cGMP compared to cAMP activated
386 gene expression up to four times higher. Both, the affinities as well as the structures of cNMP-
387 bound Clr showed no clear distinction explaining the higher degree of cGMP-dependent promoter
388 activation in *S. meliloti*. Moreover, stronger cGMP-mediated activation was not observed in *E. coli*.
389 The difference is therefore likely rooted in higher cGMP levels within the *S. meliloti* cell under the
390 *in vivo* assay conditions used in this study. Whether this is caused by endogenous AC/GCs or

differences in uptake of exogenous cNMP remains to be elucidated. Previously, it was shown that *S. meliloti* is able to produce cGMP upon overproduction of its endogenous AC/GC CyaB (23). However, which cNMP molecules are synthesized under native conditions is not yet known, as concentrations seem to be rather low. Except *cyaF2*, which encodes a putative AC or GC, all other identified genes directly regulated by Clr-cNMP code for hypothetical proteins. Eight of the Clr-cNMP regulated genes encode small proteins of up to 100 amino acids, six of these with an *N*-terminal signal peptide. *SMb20495, SMc02177* and *SMc02178* predicted to encode hypothetical proteins were previously shown to participate in autoregulation of alfalfa root hair infections (5, 30, 70). With exception of these three genes, the biological role of the other identified Clr-cNMP targets remains unclear. Because many of the direct Clr-cNMP controlled target genes encode proteins with *N*-terminal signal peptide, we speculate that most of these proteins might be relevant for extracellular functions and possibly involved in the symbiotic interaction with the host plant.

Notably, Clr-cNMP was found to directly induce transcription of the small non-coding RNA *SMelC181* of unknown function. Only a few studies reported regulation of small non-coding RNAs by a CRP. Examples are small non-coding RNAs associated to the regulation of quorum sensing, nitrogen assimilation, and galactose utilization in *E. coli*, *Salmonella*, *Vibrionales* and cyanobacteria (71-73). This makes *SmelC181* an interesting candidate for upcoming studies.

In the last two decades, cNMP signaling networks have been characterized in several bacterial species and different CRPs have been studied for their cNMP and DNA binding properties (3, 6, 10, 74). These studies imply that cNMP signaling networks and CRPs have evolved differently in the context of the different habitats of these bacteria. In this study, we showcased a CRP-like protein that has evolved to be activated both by cAMP and cGMP to perform its function in transcriptional regulation in a plant-symbiotic α-proteobacterium. The biological function of Clr bifunctionality for binding of these two cNMPs remains to be discovered. For example, it might have evolved as an adaptation to endogenous bacterial cGMP synthesis or for sensing of plant-derived cGMP. Although the *S. meliloti* Clr regulon has been extensively characterized, current knowledge of its biological functions is limited to its role in early stages of the symbiosis (5). Deciphering the

biological function of the responsiveness of Clr to cAMP and cGMP and whether subclass G CRP transcription factors are generally characterized by being activatable by cGMP or by both cGMP and cAMP is an exciting goal of future research.

MATERIALS AND METHODS

Strains, plasmids and growth conditions. Unless specified otherwise, *E. coli* was grown at 37°C in LB-Lennox medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl) with the addition of kanamycin (50 µg/mL), gentamycin (8 µg/mL or tetracycline (10 µg/mL) as needed. *S. meliloti* was grown in TY (0.5% tryptone, 0.3% yeast extract, 2.7 mM $CaCl_2$) or MOPS minimal medium (1% MOPS, 1 mM $MgSO_4$, 20 mM sodium glutamate, 20 mM mannitol, 2 mM K_2HPO_4, 250 µM $CaCl_2$, 37 µM $FeCl_3$, 4.1 µM biotin, 45.8 µM H_3BO_3, 10 µM $MnSO_4$, 1 µM $ZnSO_4$, 0.5 µM $CuSO_4$, 0.27 µM $CoCl_2$, 0.5 µM Na_2MoO_4). If required, streptomycin (600 µg/mL), kanamycin (200 µg/mL), gentamycin (30 µg/mL) or tetracycline (10 µg/mL) were added to solid medium containing agar as solidifying agent. For liquid media, the concentration of kanamycin, gentamycin or tetracycline were reduced by half.

Plasmids were transferred to *S. meliloti* by *E. coli* S17-1-mediated conjugation as previously described (75).

Promoter-probe constructs were generated by insertion of a *S. meliloti* gene upstream region of 300 to 735 bp and up to 75 bp of the associated protein or RNA coding region into the replicative low copy plasmid pPHU231-EGFP or replicative medium-copy plasmid pSRKKm-EGFP for studies in *S. meliloti* strains. For the promoters of small non-coding RNA genes, a Shine-Dalgarno sequence was included downstream of the selected promoter sequence.

Site-directed mutagenesis of putative CBSs was performed by overlap extension PCR and mutated promoter regions of the same size as the corresponding wild type promoter region were inserted into pSRKKm-EGFP.

For cloning of CBSs, two complementary oligonucleotides (1 µM each) were hybridized in T4 DNA-ligase-buffer (NEB) containing T4 polynucleotide kinase (PNK; 0.2 µl; NEB) in a total volume of 10 µl to obtain double-stranded DNA flanked by *Xba*I and *Hind*III overhangs. Phosphorylation of

the oligonucleotides was performed at 37°C for 45 min followed by a PNK deactivating step at 65°C for 20 min. The reaction was then heated at 95°C for 5 min and slowly cooled down to 30°C. 30 ng of the hybridized product was directly used as insert for ligation with pSRKKm-EGFP.

For construction of the C-terminal 3xFLAG-tagged Clr overexpression construct, the full-length *clr* coding sequence was amplified from genomic DNA of *S. meliloti* Rm2011, digested with *Nde*I and *Xba*I and inserted into plasmid pSRKKm-CF, yielding pSRKKm-*clr*-CF. Subsequently, *clr*-CF was amplified from this plasmid, digested with *Xba*I and *Hind*III and inserted into plasmid pWBT, yielding pWBT-*clr*-CF. Expression plasmids pWBT-AC/GC and pWBT-AC/GC-c*lr* are described in (23).

To obtain *S. meliloti* gene deletion mutants, gene-flanking regions of 350 to 750 bp were cloned into suicide vector pK18mobSacB. Following conjugation-mediated plasmid transfer to *S. meliloti* Rm2011 and plasmid integration into the genomic DNA by homologous recombination, transconjugants were subjected to sucrose selection as previously described (76). Gene deletions were verified by PCR. The *S. meliloti* multiple deletion mutant *cya*^{0} lacking all 28 putative class III AC/GC genes was generated by sequential gene deletions. The genomes of *cya*^{0} strains as well as the parenting strain Rm2011 were sequenced. Total DNA was purified using a DNeasy blood and tissue kit (Qiagen). DNA sequencing libraries were generated by applying the Nextera XT DNA Library Preparation kit (Illumina). Sequencing was performed on a MiSeq Desktop Sequencer (Illumina) using the MiSeq reagent kit v2, for 2 x 250 bp paired-end reads (Illumina) for Rm2011 resulting in 4.45 x 10^6 reads or the MiSeq reagent kit v3 for 2 x 75 bp paired end reads for Rm2011 *cya*^{0} resulting in 5.82 x 10^6 reads. Single-nucleotide polymorphism (SNP) detection was performed by applying CLC Genomics workbench (v10.1.1, Qiagen). Paired reads were mapped to the annotated reference genome of *S. meliloti* Rm1021 (chromosome, AL591688; pSymA, NC_003037; pSymB, AL591985). The minimum variation frequency was 50 % and at least ten reads with a minimum mismatch coverage of eight were considered. The *S. meliloti cya*^{0} strain used in this study (Fig. 4, Table S4) carries one SNP (Table S1B).

To construct the expression plasmid coding for C-terminally His_6-tagged Clr, the native encoding

sequence was amplified by PCR and inserted into *Nde*I- and *Xho*I-digested pET36b(+) (Novagen). The fusion construct was verified by DNA sequencing.

Strains, plasmids, and oligonucleotides used in this study are described in Tables S6.

EGFP fluorescence measurements. *S. meliloti* strains were grown in 100 µl TY-medium and *E. coli* BTH101 in 100 µl LB-Medium in 96-well polystyrene flat bottom plates (Greiner) at 30°C shaking at 1200 rpm. For cultures in which cNMPs were added to the medium, stationary pre-cultures of the respective strains grown in TY-medium without cNMPs were diluted 1:100 in medium containing 400 µM cAMP or cGMP, or as control in medium without cNMPs, followed by incubation for 24 h. For promoter studies in *E. coli* BTH101, stationary pre-cultures were diluted 1:100 in medium containing 100 µM IPTG followed by incubation for 16 h.

EGFP fluorescence measurements were carried out as described in (23). Relative fluorescent units (RFU) represent fluorescent values divided by the optical density (OD_{600}). The measurements were carried out in *S. meliloti* Rm2011, Rm2011 Δ*clr*, Rm2011 *cya*^{0} and *E. coli* BTH101 background. Processed data and underlying raw data of each experiment are listed in Tables S2-13.

In *S. meliloti* strains, the background fluorescence of the corresponding empty vector control was subtracted (Fig. 1 A and C, Fig. S2E, Tables S2 and S4). In measurements carried out in *E. coli* BTH101 (Fig. 3 B, Table S3), no background fluorescence was subtracted. Fluorescence fold-change of *S. meliloti* strains carrying promoter-probe constructs was calculated from the RFU of the induced strains divided by the RFU of the non-induced strains or by the RFU of the wild type divided by the RFU of the mutants (Table S2). In *E. coli*, fluorescence fold-change was calculated by the fluorescent values derived from the strains carrying pPHU231-promoter-probe-EGFP in combination with pWBT-AC/GC-*clr*-EGFP divided by the values derived from the strains carrying pPHU231-promoter-probe-EGFP in combination with pWBT-AC/GC-EGFP (Table S3). Three to four independent transconjugants and transformants of each strain containing the promoter-*egfp* constructs were used as biological replicates.

ChIP-Seq. Cultures of Rm2011 Δ*clr* carrying the *clr*-FLAG overexpression construct pWBT clr-CF were grown in 60 ml of TY or MOPS in 500 ml flasks each supplemented with either 400 µM

3',5'-cAMP or 3',5'-cGMP and 500 µM IPTG. Upon reaching an OD_{600} of 0.6, cells were fixated with 1% formaldehyde for 20 minutes at room temperature (RT). Fixation was quenched with 250 µM glycine for 20 minutes.

Cell-pellets were resuspended in 300 µl IP-buffer (50 mM HEPES-KOH pH7.8; 150 mM NaCl, 1 mM EDTA, 1% Triton X100, 0.1% sodium deoxycholate, 0.1% SDS) with 1 mM of phenylmethylsulfonyl fluoride (PMSF). Cells were lysed and DNA fragmented (~500 bp) by sonication in a Bioruptor® Plus (Diagenode) for 48 x 30 s and 30 s cooling. 300 µl sonicated product was mixed with 3 ml IP-buffer and 1 mM PMSF and centrifuged at full speed at 4°C for 30 min. 100 µl supernatant were taken and frozen at -20°C as a control. Clr-CF was immunoprecipitated with ANTI-FLAG® M2 Affinity Gel (Sigma-Aldrich). The protein was eluted with 100 µl 3x FLAG peptide elution buffer for 1 h at 4°C. Cross-links of control and ChIP-DNA were removed with 200 µM NaCl and proteins and RNA were degraded with proteinase K and RNAse A overnight at 65°C. Samples were checked in a Western blot with monoclonal anti-FLAG M2-peroxidase (HRP) antibody (1: 1000). DNA was purified with the QIAquick PCR kit. 10 ng DNA was taken for ChIP-Seq DNA library preparation and 2.5 ng for qPCR reaction. The DNA library preparation was carried out as described in (77) with the adapter sequences 2, 4-7 and 12-14 compatible with Illumina's TruSeq platform. Sequencing was performed on a MiSeq Desktop Sequencer (Illumina) using a MiSeq reagent kit v3 with 2 x 75 paired end-reads and sequence analysis was performed using CLC-Genomics workbench (v10.1.1., Qiagen). Enrichment in reads of each sample was compared to the corresponding control and mapped as peaks against the annotated genome of *S. meliloti* Rm1021. Sequences flanking each center of peak derived from the ChIP-Seq analysis were defined and summarized as FASTA sequences using the R package (78). Motifs were generated with the MEME Suite online tools MEME (v5.1) and FIMO (v5.3) (79, 29).

Electrophoretic mobility shift assay. An electrophoretic mobility shift assay (EMSA) reaction mixture contained 2 mM HEPES (pH 7.0), 30 mM NaCl, 50 mM KCl, 850 ng sonicated salmon sperm DNA (GE Healthcare), 1 µg of BSA (Sigma) and 20 ng of Cy3-labeled DNA in a final volume

of 10 µl. Native and synthetic Cy3-labeled DNA fragments were obtained by PCR with Cy3-labeled primers (Table S15D) using the corresponding pSRKKm-promoter-EGFP or pSRKKm-CBS-EGFP constructs as template, respectively. 592 bp synthetic Cy3-labeled DNA fragments were composed of the CBS flanked by pSRKKm-EGFP-derived sequence. The protein was added at 3 µg (111 µM) per reaction and 3',5'-cAMP or 3',5'-cGMP at 1 mM if indicated. The reaction mixtures were incubated at room temperature for 30 min in the dark. A mixture of 1 µl of 90% glycerol and 1.5 µl of 5x TBE buffer was added to each reaction, and 10 µl of which was loaded onto a 10% polyacrylamide gel in 1x TBE. Following electrophoresis at 9 V cm^{-1} at room temperature for 3.5 h, images were taken using a Typhoon 8600 variable mode imager (Amersham Bioscience). Sequences of every DNA fragment used in this study are listed in Table S15B and C.

Protein expression and purification. Clr was overproduced in *E. coli* BL21 (DE3) cells in LB-Lennox medium containing kanamycin (30 µg/mL) at 37°C and induced at an optical density of 0.5 (600 nm) using IPTG (0.1 mM). After 3 h after addition of 0.1 mM IPTG and harvested by centrifugation (15 min at 4°C; 4000 rpm). The pellet was resuspended in 20 mL binding buffer (50 mM sodium dihydrogen phosphate, 1000 mM sodium chloride, 10 mM imidazole, pH 7.0) and disrupted by three passages through a cold French pressure cell press at 15.2 bar cm^{-2}. The lysate was centrifuged at 18000 rpm for 60 min at 4°C and filtered through 0.45 µm membrane filter. For Clr purification the initial purification was conducted using a 5 mL Protino™ Ni-NTA Column (Macherey-Nagel). Size-exclusion chromatography over a 16/600 Superdex 200 column (GE) using elution buffer (20 mM HEPES, 300 mM NaCl, pH 7.0) was done as a polishing step. Preceding the purification, the size-exclusion buffer was optimized as described using SYPRO™ Orange dye (89).

Protein crystallisation. Clr (0.25 mM/0.44 mM) was crystallized in the presence of 25 mM cAMP/cGMP, 25 mM $MgCl_2$ and a 1.25 times excess of 19-mer/14-mer duplex target DNA (19-mer sense strand: 5'-CTA GGT AAC ATT ACT CGC-3'; 14-mer antisense strand: 5'-GCG AGT AAT GTT AC-3'). The oligonucleotide (BioCat) was prepared from single strands by heating equimolar amounts to 98°C for 5 min and slowly cooling to room temperature. The DNA oligonucleotides used

553 for the crystallisation of Clr were designed based of the genes Smc04190 and SMc00925, which
554 have previously been reported to be regulated by Clr (23).

555 Crystals of Clr in complex with cAMP were grown at 291 K by sitting-drop vapour diffusion in
556 0.09 M sodium fluoride, 0.09 M sodium bromide, 0.09 M sodium iodide, 0.1 M Tris base/BICINE
557 pH 8.5, 12.5% (v/v) MPD, 12.5% (w/v) PEG1000 and 12.5% (w/v) PEG3350. Crystals containing
558 cGMP were grown at 281 K in sitting-drop vapour diffusion plates in 0.2 M 1,6-hexanediol, 0.2 M
559 1-butanol, 0.2 M (*RS*)-1,2-butanediol, 0.2 M 2-propanol, 0.2 M 1,4-butanediol, 0.2 M
560 1,3-propanediol, 0.1 M MOPS/HEPES-Na pH 7.5, 12.5% (v/v) MPD, 12.5% (w/v) PEG1000 and
561 12.5% (w/v) PEG3350.
562 The crystals were flash-frozen in liquid nitrogen and diffraction data collected at 100 K at the
563 EMBL/DESY P13 beamline (Hamburg) using a Pilatus 6M detector (Dectris) and SLS PXI X06SA
564 beamline (Paul-Scherrer Institute, Villigen) using an Eiger 16M detector (Dectris), respectively. At
565 wavelengths of 0.976/1.0 Å the crystals diffracted to 2.8 and 3.1 Å. The data were processed using
566 XDS (81) in Space Group P21 21 21. Data reduction and scaling were done using CCP4i2
567 (v7.0.065) (82), AIMLESS in particular (Version 0.7.3). The phases were solved by molecular
568 replacement using a hybrid model derived from GlxR from *Corynebacterium glutamicum* (PDB
569 4CYD) and Crp from *Mycobacterium tuberculosis* (PDB 3MZH). The model was built using COOT
570 (v0.8.9) (83) and refinement in Phenix (v1.11.1) (84). Final refinement statistics are given in Table
571 S5. The coordinates and structure factors were deposited to the Protein Data Bank under PDB
572 codes 7PZA and 7PZB.

573 **ITC measurements.** Clr was transferred to the binding buffer (20 mM HEPES, 300 mM NaCl,
574 20 mM $MgCl_2$, pH 7.0) via a PD-10 column (Cytiva Life Sciences) and the titration calorimetry
575 measurements were performed at a Malvern MicroCAL PEAQ-ITC as previously described by (85).
576 A typical titration consisted of injecting 2 µl aliquots of 5 mM of the ligand solution into 0.3 to 0.4 mM
577 of the protein solution every 2.5 min to ensure that the titration peak returned to the baseline prior
578 to the next injection. For the measurement of DNA affinities, the setup was reversed, titrating
579 protein at 0.3 to 0.4 M containing 1 mM of cyclic nucleotide into DNA at 25 µM. The cell was

580 temperature-controlled to 25°C. Titration curves for the dilution of the ligand solution were deducted
581 from the data.

582 **HDX-MS measurements.** Sample preparation is automated with a two-arm robotic autosampler
583 (LEAP Technologies). 7.5 µL of 50 µM Clr with and without 1 mM cNMP were mixed with 67.5 µL
584 D_2O-based SEC buffer. After 10, 30, 95, 1000 and 10000 s at 25°C respectively the H/D exchange
585 was quenched by addition of equal parts quench buffer (400 mM KH_2PO_4/H_3PO_4, 2 M guanidine
586 hydrochloride, pH 2.2) at 1°C. The solution was injected into an ACQUITY UPLC M-class system
587 with HDX technology (Waters). The protein was digested online with immobilized porcine pepsin at
588 12°C at 100 µL/min (H_2O, 0.1% formic acid), and the resulting peptides were collected on a trap
589 column (2 mm × 2 cm) with POROS 20 R2 material (Thermo Scientific) at 0.5°C. After 3 min, the
590 trap column was switched online with an ACQUITY UPLC BEH C18 1.7 µm 1.0 × 100 mm column
591 (Waters), and the peptides were eluted at 0.5°C using a gradient of (A) H_2O, 0.1% formic acid and
592 (B) acetonitrile, 0.1% formic acid at 30 µL/min (5 to 35% B in 7 min, 35% to 85% B within 1 min
593 and isocratic flow 85% B for 2 min). The column was washed for 1 min at 95% B and equilibrated
594 at 5% B for 5 min after this. Peptides were ionized by electrospray ionization at 250°C source
595 capillary temperature and a spray voltage of 3.0 kV. Mass spectra were acquired on a G2-Si HDMS
596 mass spectrometer with ion mobility separation (Waters) over a range of 50 to 2,000 m/z in HDMSE
597 or HDMS mode for undeuterated and deuterated samples, respectively. Lock mass correction was
598 performed with [Glu1]-fibrinopeptide B standard (Waters). Between samples, the pepsin column
599 was washed three times with 80 µL of 4% (vol/vol) acetonitrile and 0.5 M guanidine hydrochloride,
600 and additionally, blank runs were performed between samples. Peptides were identified, and
601 deuterium uptake was determined employing the PLGS and DynamX 3.0 software suites (both
602 Waters) as described by (86).

603 **Data availability.** The ChIP-seq data have been deposited in the ArrayExpress database at
604 EMBL-EBI (www.ebi.ac.uk/arrayexpress) under the accession number E-MTAB-11788 (87). TSS
605 data of S. meliloti strain RFF625c are available from ArrayExpress (E-MTAB-12127). The Clr
606 protein crystal structure data has been deposited to the Protein Data Bank archive (PDB) under

607 accession codes 7PZA and 7PZB. All study data are included in the article and/or SI Appendix.

608 ## ACKNOWLEDGMENTS

609 The authors thank Ralf Poeschke for technical support during crystallization (MarXtal facility,
610 University of Marburg), Torsten Waldminghaus for providing the ChIP-Seq DNA-library adapter and
611 the Bioruptor Plus (Diagenode), Lotte Søgaard-Andersen for providing the Typhoon 8600 variable
612 mode imager (Amersham Bioscience), Bernadette Boomers for technical support of ChIP-seq
613 (Screening and Automation Technology facility, SYNMIKRO Technology Platform, University of
614 Marburg), Patrick Sobetzko for providing the DNA-Sequence generating R-script, and staff of the
615 beamlines P13 as operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg,
616 Germany) and X06SA of the Swiss Light Source (SLS, Villigen, Switzerland). This work was
617 supported by grants of the German Research Foundation (DFG) to LOE (ES152/14) and AB
618 (BE2121/8) and by the DFG core facility for interactions, dynamics and macromolecular assembly
619 structure. We thank Wieland Steinchen for HDX-MS data acquisition and assistance with HPLC
620 measurements.

621 ## SUPPLEMENTAL MATERIAL

622 **FIG S1** Functionality of Clr-CF.

623 **FIG S2** EMSA shift patterns of all putative Clr binding sites and promoter-probe measurements in
624 *S. meliloti.*

625 **FIG S3** Raw and integrated ITC thermograms.

626 **FIG S4** Structural features of DNA binding by Clr.

627 **TABLE S1** ChIP-seq data and analyses.

628 **TABLE S2** Clr-dependent promoter-probe activation.

629 **TABLE S3** Clr-cNMP dependent promoter-probe activation *E. coli* BTH101.

630 **TABLE S4** EGFP-derived fluorescence fold-reduction in *S. meliloti* Δclr and cya0.

631 **TABLE S5** Data collection and refinement statistics for cNMP-bound Clr-DNA complex crystal
632 structures.

633 **TABLE S6** Strains and plasmids, EMSA DNA fragments, Mutated DNA fragments,
634 oligonucleotides.

635 ## REFERENCES

636 1. J. Berthet, T. W. Rall, E. W. Sutherland, The relationship of epinephrine and glucagon to liver
637 phosphorylase. IV. Effect of epinephrine and glucagon on the reactivation of phosphorylase in
638 liver homogenates. J. Biol. Chem. 224, 463–475 (1957).

639 2. J. Berthet, E. W. Sutherland, T. W. Rall, The assay of glucagon and epinephrine with use of
640 liver homogenates. J. Biol. Chem. 229, 351–361 (1957).

641 3. S. Busby, R. H. Ebright, Transcription activation by catabolite activator protein (CAP). J. Mol.
642 Biol. 293, 199–213 (1999).

643 4. G. Soberón-Chávez, L. D. Alcaraz, E. Morales, G. Y. Ponce-Soto, L. Servín-González, The
644 transcriptional regulators of the CRP family regulate different essential bacterial functions and
645 can be inherited vertically and horizontally. Front. Microbiol. 8, 451 (2017).

646 5. C. F. Tian, A.-M. Garnerone, C. Mathieu-Demaziere, C. Masson-Boivin, J. Batut, Plant-
647 activated bacterial receptor adenylate cyclases modulate epidermal infection in the
648 *Sinorhizobium meliloti*-Medicago symbiosis. Proc. Nat. Acad. Sci. U.S.A. 109, 6751–6756
649 (2012).

650 6. J. A. Beavo, L. L. Brunton, Cyclic nucleotide research -- still expanding after half a century. Nat.
651 Rev. Mol. Cell Biol. 3, 710–718 (2002).

652 7. M. Gomelsky, cAMP, c-di-GMP, c-di-AMP and now cGMP: bacteria use them all! Mol. Microbiol.
653 79, 562–565 (2011).

654 8. G. Gosset, Z. Zhang, S. Nayyar, W. A. Cuevas, M. H. Saier, Transcriptome analysis of Crp-
655 dependent catabolite control of gene expression in *Escherichia coli*. J. Bacteriol. 186, 3516–
656 3524 (2004).

657 9. J. G. Harman, Allosteric regulation of the cAMP receptor protein. Biochim. Biophys. Acta 1547,
658 1–17 (2001).

659 10. A. Kolb, S. Busby, H. Buc, S. Garges, S. Adhya, Transcriptional regulation by cAMP and its
660 receptor protein. Annu. Rev. Biochem. 62, 749–797 (1993).

661 11. H.-S. Won, Y.-S. Lee, S.-H. Lee, B.-J. Lee, Halil. Biochim. Biophys. Acta 1794, 1299–1308
662 (2009).

663 12. T. A. Kohl, A. Tauch, The GlxR regulon of the amino acid producer *Corynebacterium*
664 *glutamicum*: Detection of the corynebacterial core regulon and integration into the
665 transcriptional regulatory network model. J. Biotechnol. 143, 239–246 (2009).

666 13. G. S. Knapp, K. A. McDonough, Cyclic AMP signaling in Mycobacteria. Microbiol. Spectr. 2,
667 doi.org/10.1128/microbiolspec.MGM2-0011-2 (2014).

668 14. C. Kahramanoglou, T. Cortes, N. Matange, D. M. Hunt, S. S. Visweswariah, D. B. Young, R. S.
669 Buxton, Genomic mapping of cAMP receptor protein (CRP Mt) in *Mycobacterium tuberculosis*:

Relation to transcriptional start sites and the role of CRP Mt as a transcription factor. Nucl. Acids Res. 42, 8320–8329 (2014).

15. B. Jungwirth, C. Sala, T. A. Kohl, S. Uplekar, J. Baumbach, S. T. Cole, A. Pühler, A. Tauch, High-resolution detection of DNA binding sites of the global transcriptional regulator GlxR in *Corynebacterium glutamicum*. Microbiology 159, 12–22 (2013).

16. H. Körner, H. J. Sofia, W. G. Zumft, Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: Exploiting the metabolic spectrum by controlling alternative gene programs. FEMS Microbiol. Rev. 27, 559–592 (2003).

17. J. B. Shabb, J. D. Corbin, Cyclic nucleotide-binding domains in proteins having diverse functions. J. Biol. Chem. 267, 5723–5726 (1992).

18. H. Youn, R. L. Kerby, J. Koh, G. P. Roberts, A C-helix residue, Arg-123, has important roles in both the active and inactive forms of the cAMP receptor protein. J. Biol. Chem. 282, 3632–3639 (2007).

19. C. L. Lawson, D. Swigon, K. S. Murakami, S. A. Darst, H. M. Berman, R. H. Ebright, Catabolite activator protein: DNA binding and transcription activation. Curr. Opin. Struct. Biol. 14, 10–20 (2004).

20. B. Benoff, C. L. Lawson, G. Parkinson, J. Liu, E. Blatter, Y. W. Ebright, H. M. Berman, R. H. Ebright, Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. Science 297, 1562–1566 (2002).

21. D. B. McKay, T. A. Steitz, Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature 290, 744–749 (1981).

22. S. Roychowdhury, Q. Dong, C. E. Bauer, DNA-binding properties of a cGMP-binding CRP homologue that controls development of metabolically dormant cysts of *Rhodospirillum centenum*. Microbiology 161, 2256–2264 (2015).

23. E. Krol, C. Klaner, P. Gnau, V. Kaever, L.-O. Essen, A. Becker, Cyclic mononucleotide- and Clr-dependent gene regulation in *Sinorhizobium meliloti*. Microbiology 162, 1840–1856 (2016).

24. R. H. Ebright, S. F. Le Grice, J. P. Miller, J. S. Krakow, Analogs of cyclic AMP that elicit the biochemically defined conformational change in catabolite gene activator protein (CAP) but do not stimulate binding to DNA. J. Mol. Biol. 182, 91–107 (1985).

25. B. Gourion, F. Berrabah, P. Ratet, G. Stacey, Rhizobium–legume symbioses: The crucial role of plant immunity. Trends Plant Sci. 20, 186–194 (2015).

26. K. M. Jones, H. Kobayashi, B. W. Davies, M. E. Taga, G. C. Walker, How rhizobial symbionts invade plants: The *Sinorhizobium–Medicago* model. Nat. Rev. Microbiol. 5, 619–633 (2007).

27. J. Batut, M. L. Daveran-Mingot, M. David, J. Jacobs, A. M. Garnerone, D. Kahn, *fixK*, a gene homologous with *fnr* and *crp* from *Escherichia coli*, regulates nitrogen fixation genes both positively and negatively in *Rhizobium meliloti*. EMBO J. 8, 1279–1286 (1989).

28. F. Galibert, T. M. Finan, S. R. Long, A. Pühler, P. Abola, F. Ampe et al., The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293, 668–672 (2001).

29. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018 (2011).

30. L. Zou, A. Gastebois, C. Mathieu-Demazière, F. Sorroche, C. Masson-Boivin, J. Batut, A.-M. Garnerone, Transcriptomic insight in the control of legume root secondary infection by the *Sinorhizobium meliloti* transcriptional regulator Clr. Front. Microbiol. 8, 292 (2017).

31. J.-P. Schlüter, J. Reinkensmeier, S. Daschkey, E. Evguenieva-Hackenberg, S. Janssen, S. Jänicke, J. D. Becker, R. Giegerich, A. Becker, A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. BMC Genomics 11, 245 (2010).

32. J.-P. Schlüter, J. Reinkensmeier, M. J. Barnett, C. Lang, E. Krol, R. Giegerich, S. R. Long, A. Becker, Global mapping of transcription start sites and promoter motifs in the symbiotic α-proteobacterium *Sinorhizobium meliloti* 1021. BMC Genomics 14, 156 (2013).

33. E. Sallet, B. Roux, L. Sauviac, M.-F. Jardinaud, S. Carrère, T. Faraut, F. de Carvalho-Niebel, J. Gouzy, P. Gamas, D. Capela, C. Bruand, T. Schiex, Next-generation annotation of prokaryotic genomes with EuGene-P: application to *Sinorhizobium meliloti* 2011. DNA Res. 20, 339–354 (2013).

34. Y. Zhou, T. J. Merkel, R. H. Ebright, Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP) II. Role at class I and class II CAP-dependent promoters. J. Mol. Biol. 243, 603–610 (1994).

35. C. Ushida, H. Aiba, Helical phase dependent action of CRP: effect of the distance between the CRP site and the -35 region on promoter activity. Nucleic Acids Res 18, 6325–6330 (1990).

36. R. H. Ebright, Transcription activation at Class I CAP-dependent promoters. Mol. Microbiol. 8, 797–802 (1993).

37. J. Małecki, A. Polit, Z. Wasylewski, Kinetic studies of cAMP-induced allosteric changes in cyclic AMP receptor protein from *Escherichia coli*. J. Biol. Chem. 275, 8480–8486 (2000).

38. M. Stapleton, I. Haq, D. M. Hunt, K. B. Arnvig, P. J. Artymiuk, R. S. Buxton, J. Green, Mycobacterium tuberculosis cAMP receptor protein (Rv3676) differs from the *Escherichia coli* paradigm in its cAMP binding and DNA binding properties and transcription activation properties. J. Biol. Chem. 285, 7016–7027 (2010).

39. Y. Tutar, Neglected role of cAMP receptor protein monomer. Mol. Biol. Rep. 39, 4261–4265 (2012).

40. S.-P. Scott, S. Jarjous, Proposed structural mechanism of *Escherichia coli* cAMP receptor protein cAMP-dependent proteolytic cleavage protection and selective and nonselective DNA binding. Biochemistry 44, 8730–8748 (2005).

41. M. C. M. Reddy, S. K. Palaninathan, J. B. Bruning, C. Thurman, D. Smith, J. C. Sacchettini, Structural insights into the mechanism of the allosteric transitions of *Mycobacterium tuberculosis* cAMP receptor protein. J. Biol. Chem. 284, 36581–36591 (2009).

42. W. Epstein, L. B. Rothman-Denes, J. Hesse, Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. Proc. Nat. Acad. of Sci. U.S.A. 72, 2300–2304 (1975).

43. B. Aykaç Fas, Y. Tutar, T. Haliloğlu, Dynamic fluctuations provide the basis of a conformational switch mechanism in apo cyclic AMP receptor protein. PLoS Comput. Biol. 9, e1003141 (2013).

750 44. N. Popovych, S.-R. Tzeng, M. Tonelli, R. H. Ebright, C. G. Kalodimos, Structural basis for
751 cAMP-mediated allosteric control of the catabolite activator protein. Proc. Nat. Acad. of Sci.
752 U.S.A. 106, 6927–6932 (2009).

753 45. R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, R. S. Mann, Origins of specificity in protein-
754 DNA recognition. Ann. Rev. Biochem. 79, 233–269 (2010).

755 46. R. D. Finn, T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H.-Y. Chang, Z.
756 Dosztányi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G. L. Holliday, H. Huang, X. Huang, I.
757 Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D. A. Natale, M. Necci, G. Nuka,
758 C. A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S. C. Potter, N. D. Rawlings, N. Redaschi, L.
759 Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato,
760 G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, I. Xenarios, L.-S. Yeh, S.-Y.
761 Young, A. L. Mitchell, InterPro in 2017-beyond protein family and domain annotations. Nucl.
762 Acids Res. 45, D190–D199 (2017).

763 47. S. C. Schultz, G. C. Shields, T. A. Steitz, Crystal structure of a CAP-DNA complex: The DNA
764 is bent by 90 degrees. Science 253, 1001–1007 (1991).

765 48. P. D. Townsend, B. Jungwirth, F. Pojer, M. Bußmann, V. A. Money, S. T. Cole, A. Pühler, A.
766 Tauch, M. Bott, M. J. Cann, E. Pohl, The crystal structures of apo and cAMP-bound GlxR from
767 *Corynebacterium glutamicum* reveal structural and dynamic changes upon cAMP binding in
768 CRP/FNR family transcription factors. PloS One 9, e113265 (2014).

769 49. Y. Akhter, M. Wilmanns, Crystal structure of cAMP receptor protein from *Mycobacterium*
770 *tuberculosis* in complex with cAMP and its DNA binding element (2011). PDB DOI:
771 10.2210/pdb3MZH/pdb

772 50. S.-H. Seok, H. Im, H.-S. Won, M.-D. Seo, Y.-S. Lee, H.-J. Yoon, M.-J. Cha, J.-Y. Park, B.-J.
773 Lee, Structures of inactive CRP species reveal the atomic details of the allosteric transition that
774 discriminates cyclic nucleotide second messengers. Acta Crystallogr. D. Biol. Crystallogr. 70,
775 1726–1742 (2014).

776 51. D.-W. Sim, J. W. Choi, J.-H. Kim, K.-S. Ryu, M. Kim, H.-W. Yu, K.-S. Jo, E.-H. Kim, M.-D. Seo,
777 Y. H. Jeon, B.-J. Lee, Y. P. Kim, H.-S. Won, C-terminal dimerization of apo-cyclic AMP receptor
778 protein validated in solution. FEBS Lett. 591, 1064–1070 (2017).

779 52. H. S. Won, T. Yamazaki, T. W. Lee, M. K. Yoon, S. H. Park, Y. Kyogoku, B. J. Lee, Structural
780 understanding of the allosteric conformational change of cyclic AMP receptor protein by cyclic
781 AMP binding. Biochemistry 39, 13953–13962 (2000).

782 53. Y. Tutar, Chemical linkage at allosteric activation of *E. coli* cAMP receptor protein. Prot. J. 27,
783 21–29 (2008).

784 54. S. Harteis, S. Schneider, Making the bend: DNA tertiary structure and protein-DNA interactions.
785 Int. J. Mol. Sci. 15, 12335–12363 (2014).

786 55. J. Dai, S.-H. Lin, C. Kemmis, A. J. Chin, J. C. Lee, Interplay between site-specific mutations
787 and cyclic nucleotides in modulating DNA recognition by *Escherichia coli* cyclic AMP receptor
788 protein. Biochemistry 43, 8901–8910 (2004).

789 56. E. Fic, P. Bonarek, A. Gorecki, S. Kedracka-Krok, J. Mikolajczak, A. Polit, M. Tworzydlo, M.
790 Dziedzicka-Wasylewska, Z. Wasylewski, cAMP receptor protein from *Escherichia coli* as a
791 model of signal transduction in proteins--a review. J. Mol. Microbiol. Biotechnol. 17, 1–11
792 (2009).

793 57. G. Parkinson, C. Wilson, A. Gunasekera, Y. W. Ebright, R. H. Ebright, R. E. Ebright, H. M.
794 Berman, Structure of the CAP-DNA complex at 2.5 angstroms resolution: A complete picture
795 of the protein-DNA interface. J. Mol. Biol. 260, 395–408 (1996).

796 58. B. P. Hudson, J. Quispe, S. Lara-González, Y. Kim, H. M. Berman, E. Arnold, R. H. Ebright, C.
797 L. Lawson, Three-dimensional EM structure of an intact activator-dependent transcription
798 initiation complex. Proc. Nat. Acad. of Sci. U.S.A. 106, 19830–19835 (2009).

799 59. B. Liu, C. Hong, R. K. Huang, Z. Yu, T. A. Steitz, Structural basis of bacterial transcription
800 activation. Science 358, 947–951 (2017).

801 60. Y. Feng, Y. Zhang, R. H. Ebright, Structural basis of transcription activation. Science 352,
802 1330–1333 (2016).

803 61. W. Shi, Y. Jiang, Y. Deng, Z. Dong, B. Liu, Visualization of two architectures in class-II CAP-
804 dependent transcription activation. PLoS Biol. 18, e3000706 (2020).

805 62. H. Youn, J. Koh, G. P. Roberts, Two-state allosteric modeling suggests protein equilibrium as
806 an integral component for cyclic AMP (cAMP) specificity in the cAMP receptor protein of
807 *Escherichia coli*. J. Bacteriol. 190, 4532–4540 (2008).

808 63. I. T. Weber, T. A. Steitz, Structure of a complex of catabolite gene activator protein and cyclic
809 AMP refined at 2.5 Å resolution. J. Mol. Biol. 198, 311–326 (1987).

810 64. S.-R. Tzeng, C. G. Kalodimos, Dynamic activation of an allosteric regulatory protein. Nature
811 462, 368–372 (2009).

812 65. H. Youn, R. L. Kerby, M. Conrad, G. P. Roberts, Study of highly constitutively active mutants
813 suggests how cAMP activates cAMP receptor protein. J. Biol. Chem. 281, 1119–1127 (2006).

814 66. S. M. Gunasekara, M. N. Hicks, J. Park, C. L. Brooks, J. Serate, C. V. Saunders, S. K. Grover,
815 J. J. Goto, J.-W. Lee, H. Youn, Directed evolution of the *Escherichia coli* cAMP receptor protein
816 at the cAMP pocket. J. Biol. Chem. 290, 26587–26596 (2015).

817 67. L. Bracco, D. Kotlarz, A. Kolb, S. Diekmann, H. Buc, Synthetic curved DNA sequences can act
818 as transcriptional activators in *Escherichia coli*. EMBO J. 8, 4289–4296 (1989).

819 68. E. A. Pyles, J. C. Lee, Escherichia coli cAMP receptor protein-DNA complexes. 2. Structural
820 asymmetry of DNA bending. Biochemistry 37, 5201–5210 (1998).

821 69. S.-H. Lin, J. C. Lee, Determinants of DNA bending in the DNA−cyclic AMP receptor protein
822 complexes in Escherichia coli. Biochemistry 42, 4809–4818 (2003).

823 70. A.-M. Garnerone, F. Sorroche, L. Zou, C. Mathieu-Demazière, C. F. Tian, C. Masson-Boivin, J.
824 Batut, NsrA, a predicted β-barrel outer membrane protein involved in plant signal perception
825 and the control of secondary infection in *Sinorhizobium meliloti*. J. Bacteriol. 200, e00019-18
826 (2018).

827 71. K. Papenfort, V. Pfeiffer, S. Lucchini, A. Sonawane, J. C. D. Hinton, J. Vogel, Systematic
828 deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent
829 riboregulator of OmpX synthesis. Mol. Microbiol. 68, 890–906 (2008).

830 72. N. de Lay, S. Gottesman, The Crp-activated small noncoding regulatory RNA CyaR (RyeE)
831 links nutritional status to group behavior. J. Bacteriol. 191, 461–476 (2009).

73. C. Bækkedal, P. Haugen, The Spot 42 RNA: A regulatory small RNA with roles in the central metabolism. RNA Biol. 12, 1071–1077 (2015).

74. K. A. McDonough, A. Rodriguez, The myriad roles of cyclic AMP in microbial pathogens: from signal to sword. Nat. Rev. Microbiol. 10, 27–38 (2011).

75. E. Krol, A. Becker, Rhizobial homologs of the fatty acid transporter FadL facilitate perception of long-chain acyl-homoserine lactone signals. Proc. Nat. Acad. Sci. U.S.A. 111, 10702–10707 (2014).

76. A. Schäfer, A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach, A. Pühler, Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145, 69–73 (1994).

77. E. Ford, C. Nikopoulou, A. Kokkalis, D. Thanos, A method for generating highly multiplexed ChIP-seq libraries. BMC Res. Notes 7, 312 (2014).

78. R Core Team, R: A Language and Environment for Statistical Computing.

79. T. L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings. International Conference on Intelligent Systems for Molecular Biology 2, 28–36 (1994).

80. U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, P. Nordlund, Thermofluor-based high-throughput stability optimization of proteins for structural studies. Analyt. Biochem. 357, 289–298 (2006).

81. W. Kabsch, XDS. *Acta crystallographica. Section D, Biol. Crystallogr.* 66, 125–132 (2010).

82. E L. Potterton, J. Agirre, C. Ballard, K. Cowtan, E. Dodson, P. R. Evans, H. T. Jenkins, R. Keegan, E. Krissinel, K. Stevenson, A. Lebedev, S. J. McNicholas, R. A. Nicholls, M. Noble, N. S. Pannu, C. Roth, G. Sheldrick, P. Skubak, J. Turkenburg, V. Uski, F. von Delft, D. Waterman, K. Wilson, M. Winn, M. Wojdyr, CCP4i2: The new graphical user interface to the CCP4 program suite. Acta crystallographica. Section D, Struct. Biol. 74, 68–84 (2018).

83. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. Acta crystallographica. Section D, Biol. Crystallogr. 66, 486–501 (2010).

84. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. Acta crystallographica. Section D, Struct. Biol. 75, 861–877 (2019).

85. T. Wiseman, S. Williston, J. F. Brandts, L.-N. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Analyt. Biochem. 179, 131–137 (1989).

86. M. Osorio-Valeriano, F. Altegoer, W. Steinchen, S. Urban, Y. Liu, G. Bange, M. Thanbichler, ParB-type DNA segregation proteins are CTP-dependent molecular switches. Cell 179, 1512-1524.e15 (2019).

87. A. Athar, A. Füllgrabe, N. George, H. Iqbal, L. Huerta, A. Ali, C. Snow, N. A. Fonseca, R. Petryszak, I. Papatheodorou, U. Sarkans, A. Brazma, ArrayExpress update - from bulk to single-cell expression data. Nucl. Acids Res. 47, D711–D715 (2019).

874 88. C. Lang, M. J. Barnett, R. F. Fisher, L. S. Smith, M. E. Diodati, S. R. Long, Most *Sinorhizobium*
875 *meliloti* extracytoplasmic function sigma factors control accessory functions. mSphere 3,
876 e00454-18 (2018).

FIGURES AND TABLES

TABLE 1 Isothermal titration calorimetry measurements for the complex assembly. The measurements were conducted at 25°C. The corresponding raw data can be found in Fig. S3.

Cell	Syringe	K_D (μM)	n	ΔH (kcal/mol)	ΔG (kcal/mol)	-TΔS (kcal/mol)
Clr	cAMP	5.8 ± 1.4	1.02 ± 0.016	-2.22 ± 0.05	-7.16 ± 0.12	-4.94 ± 0.22
Clr, DNA	cAMP	6.7 ± 2.7	0.57 ± 0.019	-11.5 ± 0.77	-7.08 ± 0.16	4.42 ± 0.36
DNA	Clr, cAMP	6.0 ± 4.7	0.39 ± 0.201	-11.0 ± 7.30	-7.13 ± 0.10	3.87 ± 1.37
Clr	cGMP	23.8 ± 14.5	1.49 ± 0.081	0.29 ± 0.02	-6.31 ± 0.05	-6.60 ± 0.07
Clr, DNA	cGMP	10.7 ± 3.9	0.57 ± 0.022	-4.91 ± 0.02	-6.79 ± 0.09	-1.88 ± 0.19
DNA	Clr, cGMP	4.5 ± 12.4	0.45 ± 0.408	-3.47 ± 5.33	-7.30 ± 0.02	-3.83 ± 0.02

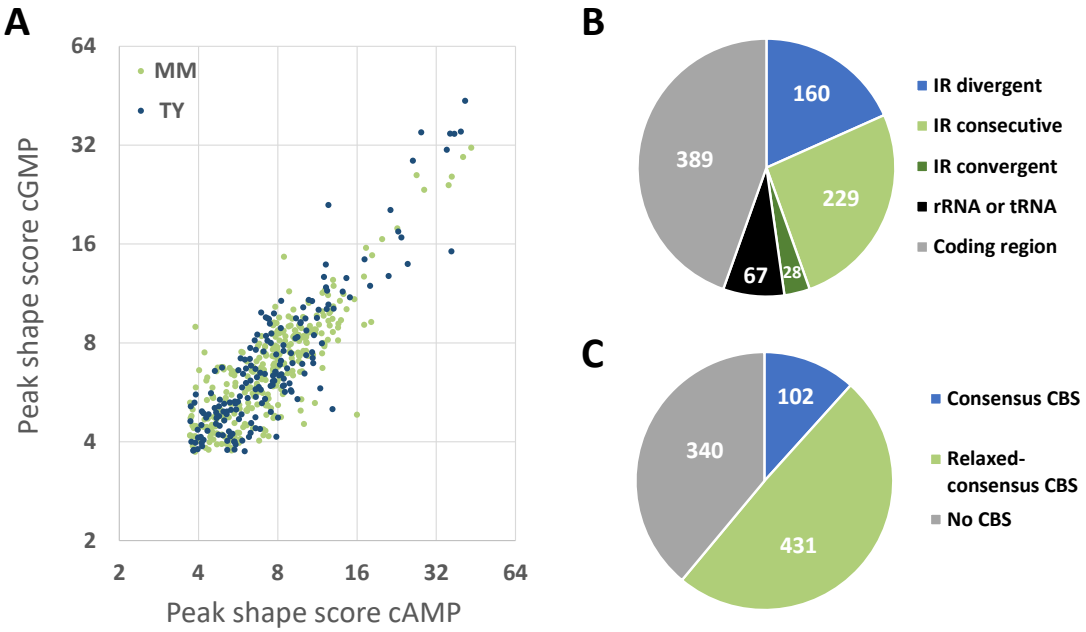


FIG 1 ChIP-seq-assisted screening for Clr-cAMP and Clr-cGMP binding sites. (A) Correlation between peak shape score values of ChIP-seq peaks detected in samples of cells grown in TY or MM, both supplemented with either cAMP or cGMP. (B) Location of Clr ChIP-seq peaks within genomic features. IR, intergenic region; IR divergent, IR between two genes transcribed in divergent directions; IR consecutive, IR between two genes transcribed in the same direction; IR convergent, IR between two genes transcribed in converging directions; rRNA or tRNA, rRNA or tRNA encoding region; coding region, protein coding region. (C) Presence of putative Clr binding sites (CBSs) in the vicinity of ChIP-seq peaks. Consensus CBS: GTNNCNNNNGNNAC. Relaxed consensus CBS, one mismatch to the consensus was allowed.

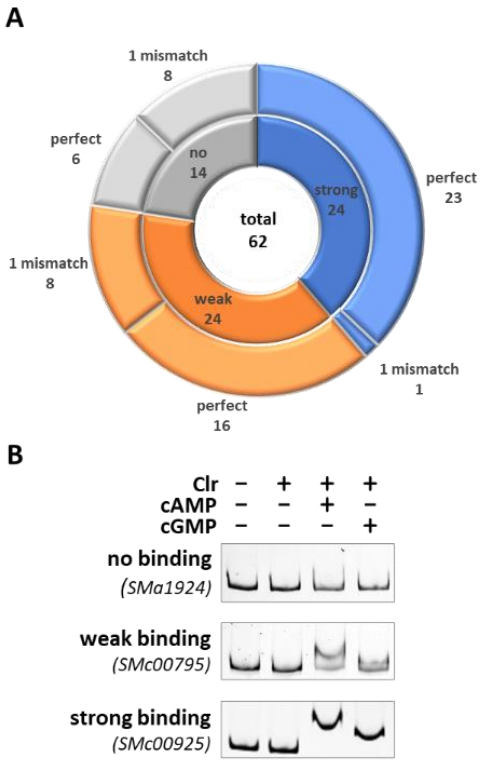


FIG 2 EMSA-based subclassification of CBS. All DNA fragments used for EMSAs contained at least one putative CBS matching the consensus binding site sequence either perfectly or with one mismatch. (A) EMSAs were subclassified into strong (blue), weak (orange) and no binding (grey). Numbers of perfect and imperfect binding sites are indicated for each subclass. (B) Representative EMSAs are shown for each subclass. These were performed with upstream regions of *SMa1924* (no binding), *SMc00795* (weak binding) and *SMc00925* (strong binding), which each contained a perfect match binding site.

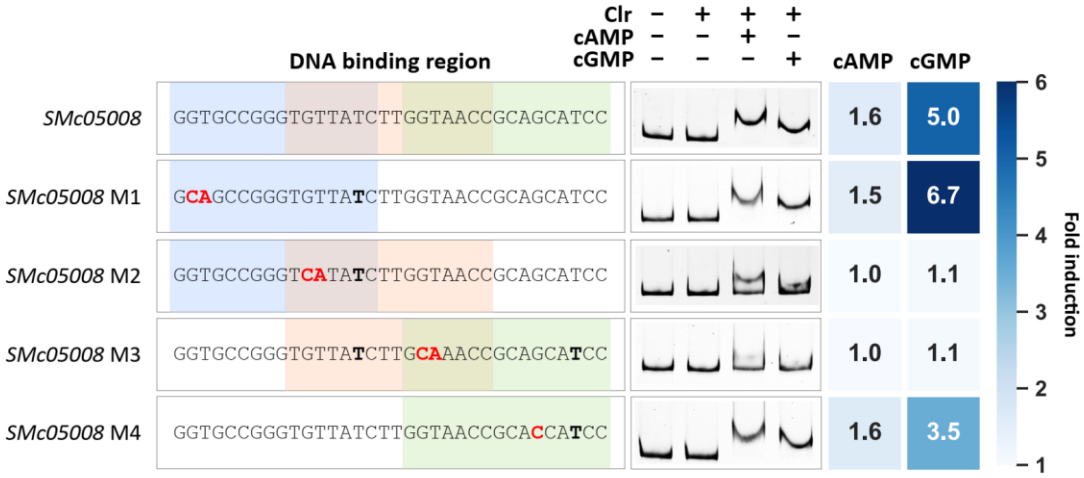


FIG 3 DNA binding of Clr-cAMP and Clr-cGMP at the *SMc05008* promoter. The promoter region of *SMc05008* has three possible DNA binding sites (colored), with one mismatch each. Mutations indicated in red were introduced at four different positions (M1-4). Nucleotides not matching the consensus Clr binding sequence are printed in bold. The heatmap shows the cAMP- or cGMP-induced fold change of EGFP-mediated fluorescence compared to non-inducing conditions (no cNMP added).

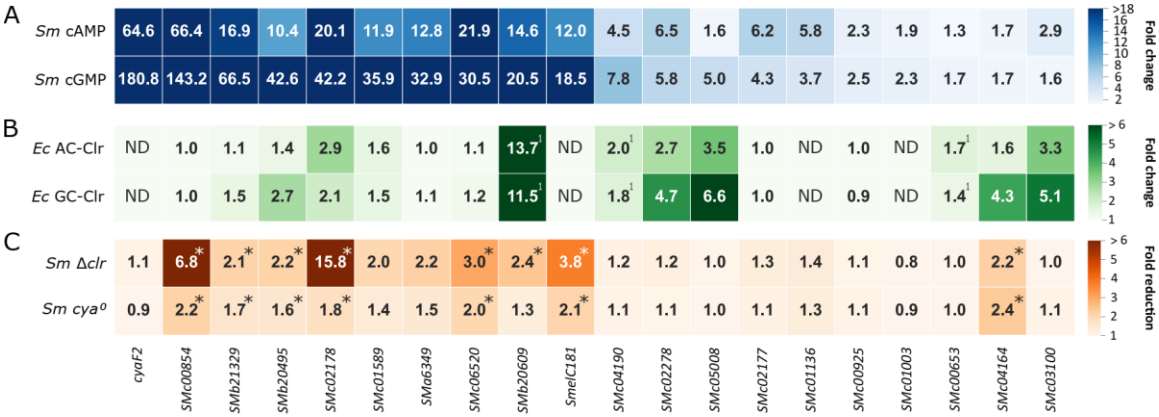


FIG 4 Promoter-probe measurements in different hosts. (A) Fold-change of EGFP fluorescence signal mediated by promoter-probe constructs in *S. meliloti* Rm2011 in cultures supplemented with cAMP or cGMP. Ratios derive from relative fluorescent units (RFU) of cNMP-induced against uninduced cultures. (B) Fold-change of EGFP fluorescence signal mediated by promoter-probe constructs in *E. coli* BTH101 upon induced production of an AC (CyaG1 of *S. meliloti*) or a GC (Cya2 of *Synechocystis* sp.) together with Clr (*S. meliloti*) in LB. Ratios shown derive from fluorescence units of strains under AC/GC-Clr production conditions compared to control strains producing the AC or GC but not Clr. [1] Values derived from (23). ND indicates data not determined. (C) Fold-reduction of EGFP fluorescence signal mediated by promoter-probe constructs in *S. meliloti* Rm2011 Δ*clr* and Rm2011 *cya*^{0} compared to Rm2011 wild type. *Indicate statistically significant fluorescent reduction ($p < 0.05$) in combination with a wild type RFU threshold of ≥ 80. Values of each biological replicate and standard deviation of samples measured in A-C are given in Tables S2-13.

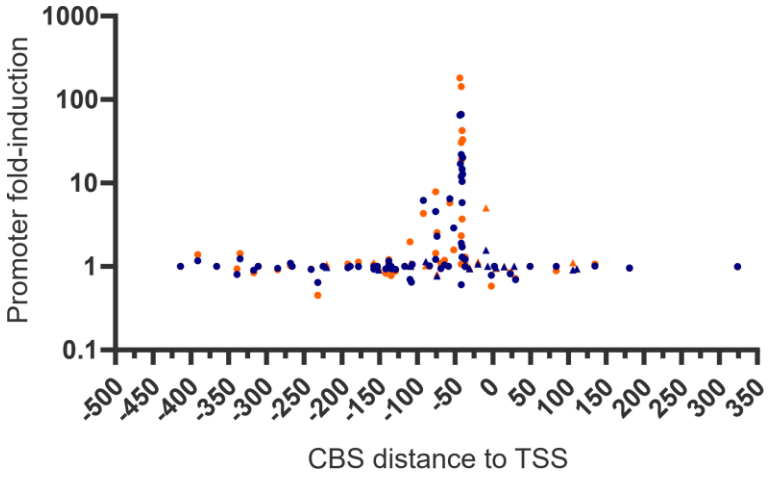


FIG 5 Promoter-probe induction relative to CBS location. All promoter regions with TSS data available and containing a putative CBS with a perfect (circle) or imperfect (one mismatch) (triangle) consensus sequence are displayed. TSS were derived from RNA-seq data of *S. meliloti* wild type strains (32, 33) and of the *ecf*⁻ *S. meliloti* strain RFF625c (88). Promoter induction ratios derived from cultures supplemented with cAMP (blue) or cGMP (orange) are shown in Fig. S2E and Table S2.

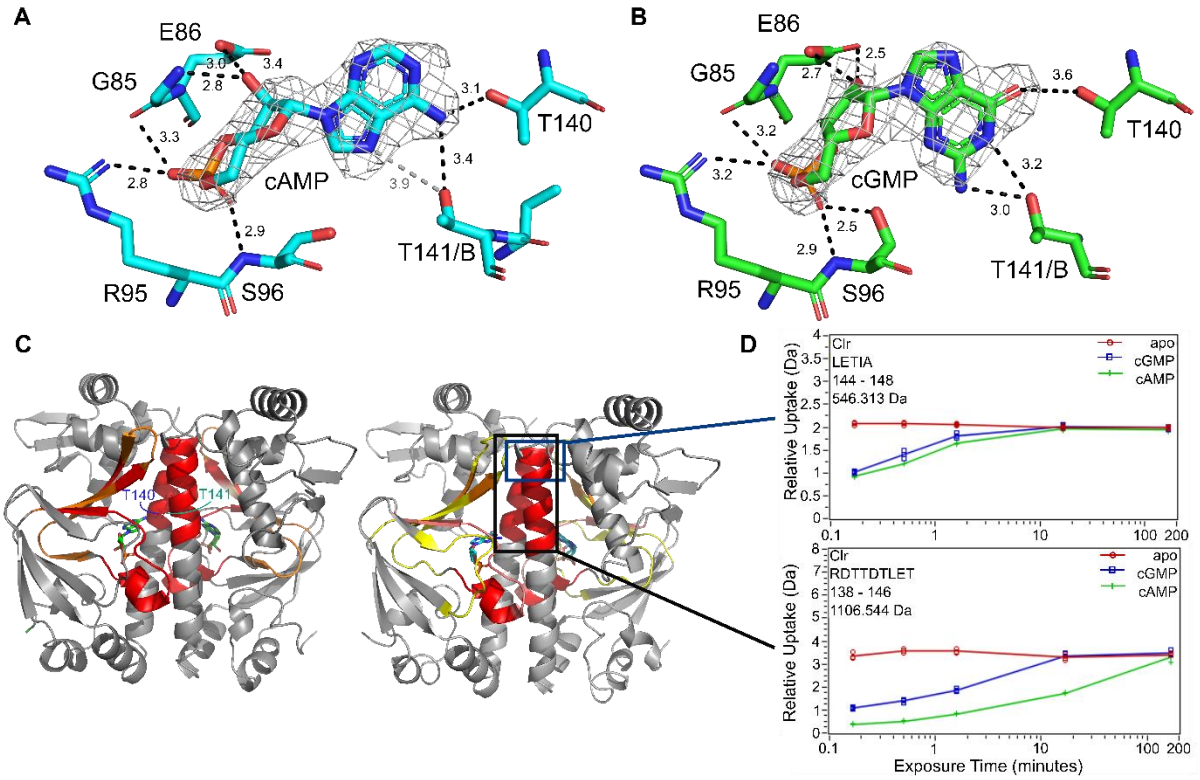
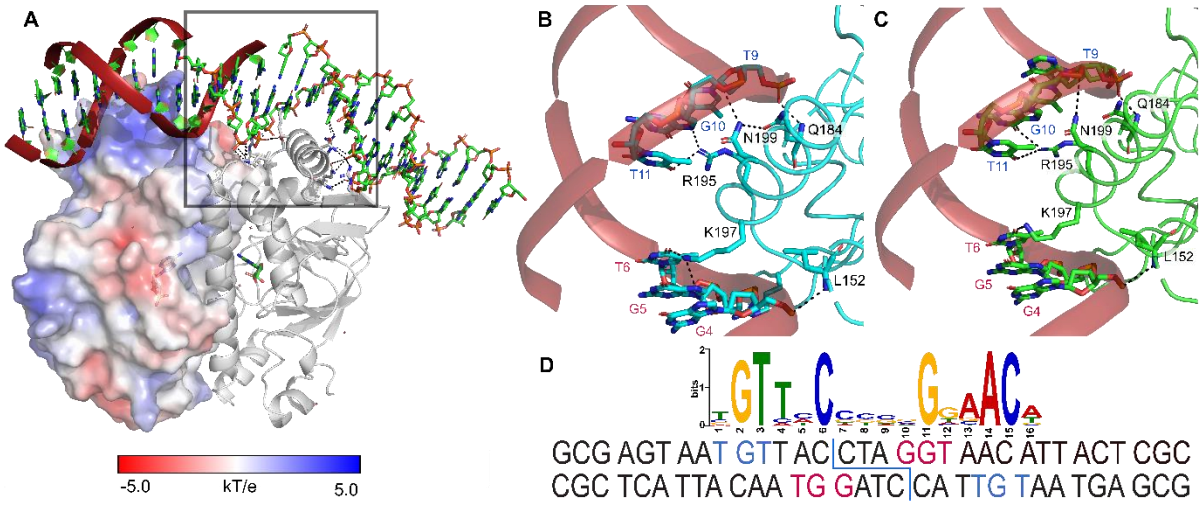


FIG 6 Characteristics of the Clr-ligand interaction. (A, B) Nucleotide binding environment in the cAMP- and cGMP-bound crystal structures. The sugar and phosphate moiety of the ligand interact with G85 and E86 of helix α4, as well as R95 and S96. The sidechain of Ser96 is tilted towards the ligand in the cAMP-bound structure and Thr141 of the adjacent subunit is in closer proximity to it.; The primary amine of the adenosine is coordinating Thr140 from α5 in the same subunit and Thr141 from the adjacent one. The ligand 2 mF_o−DF_c experimental electron density maps contoured at 1.3 σ are shown in grey. Binding interactions are indicated in Å. (C) Areas of reduced relative hydrogen-deuterium exchange upon binding of cAMP (left) and cGMP (right) coloured in from 5-20% in 5% increments (yellow, orange, red). In addition to the expected displacement of solvent in the binding cavity, the *C*-terminal half of helix 5 (black box) shows a reduction of exchange, that can be attributed to a coil-to-helix transformation. (D) Time-dependent deuterium exchange graphs for two key peptides within helix 5 show, that the shielding cannot be solely due to interaction with T140 and T141, but is the result of a change in secondary structure. Peptides not including both exemplified by ^{144}LETIA (blue box) also show a decrease in H/D exchange.



948
949 **FIG 7** Interaction of Clr with its DNA recognition site. (A) The surface of the helix-turn-helix DNA
950 binding domain is positively charged to interact with the phosphate backbone. Helix α8 is inserted
951 into the major groove and provides base readout via direct hydrogen bonding interactions. The
952 double strand break above the subunit interface facilitates the DNA kinking of about 40°. (B, C)
953 Direct interactions of Clr with the DNA major groove upon cAMP (B) and cGMP (C) binding. The
954 DNA bases involved in the interaction are highly conserved. (D) The synthetic oligonucleotide used
955 in crystallisation has a double strand break with sticky ends right above the subunit interface; the
956 5'-oligonucleotides of the annealed core motif (GCGA…) belong to chains D and F, the 3'-
957 oligonucleotides to chains C and E of the structure. The design was based on the consensus motif
958 GTNNCNNNNGNNAC.