The EMBO Journal Vol.16 No.8 pp.2108–2118, 1997

Protein-induced fit: the CRP activator protein changes sequence-specific DNA recognition by the CytR repressor, a highly flexible LacI member

Henrik Pedersen[1] and Poul Valentin-Hansen[2]

Department of Molecular Biology, Odense University, Campusvej 55, DK-5230 Odense M, Denmark

[1]Present address: Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA

[2]Corresponding author

The CytR repressor and the cAMP receptor protein (CRP) bind cooperatively to several promoters in *Escherichia coli* to repress transcription initiation. The synergistic binding is mediated by protein–protein interactions between the two regulators. Here, *in vitro* selection experiments have been used to examine the DNA-binding characteristics of CytR, by itself and when co-binding with cAMP–CRP. We show that the optimal CytR-binding site consists of two octamer repeats, in direct or inverted orientation, and separated by 2 bp. However, when co-binding with cAMP–CRP, CytR instead recognizes inverted repeats separated by 10–13 bp, or direct repeats separated by 1 bp. The configurations of the latter set of operators correlate well with the configurations of natural CytR targets. Thus, cAMP–CRP induces conformational changes in CytR so that the repressor fits the natural targets. Most strikingly, CytR can adopt widely different conformations that are equally favored energetically for complex formation with cAMP–CRP. We propose that this structural adaptability is essential for CytR repression of promoters with diverse architectures. We discuss these novel concepts in the context of the CRP/CytR regulatory system, as well as the structural and functional implications for multiprotein–DNA complex formation in general.

Keywords: cAMP-CRP/CytR/*in vitro* selection/protein–DNA interactions/protein–protein interactions

Introduction

The selective regulation of cellular processes such as site-specific recombination, transcription and DNA replication depends upon the recognition of specific DNA sites by DNA-binding proteins. In the simplest cases, the DNA-binding domain of the protein carries all the information necessary to specify its site of action. There are, however, many systems in which additional factors are required to specify the exact binding site of a protein. The complexity of these systems is often very high, and the molecular mechanisms that provide affinity and specificity remain elusive.

In principle, a DNA-binding protein may improve the DNA-binding specificity of another protein (i) by providing additional contacts through its surface, (ii) by creating a DNA conformation that is better recognized by the partner protein, or (iii) by inducing a conformational change in the partner protein that promotes its interaction with the operator. Here, we have sought the mechanism by which sequence-specific DNA recognition is achieved in a relatively simple prokaryotic system. In this system, the global activator protein cAMP receptor protein (CRP; also referred to as the catabolite gene activator protein, CAP), guides another DNA-binding protein, the CytR repressor, to a number of binding sites that share minimal sequence homology.

The CytR repressor belongs to the LacI family of regulators, and possesses, like the other members, an N-terminal helix–turn–helix (HTH) DNA-binding motif (Valentin-Hansen *et al.*, 1986; Weickert and Adhya, 1992). However, unlike a typical bacterial repressor, CytR binds with only modest affinity to its operators and cannot repress its cognate promoters independently *in vivo*. This deficiency is overcome by interaction with DNA-bound cAMP–CRP complexes. Thus, repression involves the formation of nucleoprotein complexes held together by multiple protein–DNA and protein–protein interactions (for review, see Valentin-Hansen *et al.*, 1996). The synergy in the system is prominent; depending on which promoter is examined, the cAMP–CRP activator complex strengthens binding of CytR from 100- to several thousand-fold *in vitro* (Gerlach *et al.*, 1991; Pedersen *et al.*, 1991, 1992, 1995; Holst *et al.*, 1992). At physiological CytR concentrations, the CytR–DNA interaction is absolutely required for formation of repression complexes. However, when overexpressed, a mutant CytR protein lacking its DNA-binding domain can repress the *deoP2* promoter *in vivo* (Søgaard-Andersen and Valentin-Hansen, 1993). Thus, CytR's site of action can be specified solely by the contact with CRP. The apparently low sequence homology among the natural CytR-binding sites has, therefore, been explained by the fact that interactions between neighboring proteins on the DNA could provide the repression complex with adequate specificity, even without very specific CytR–DNA interactions.

Recently, the DNA recognition specificities of several transcription factors including HSF, SRF, MyoD, E2A, Oct1 and GCN4 have been analyzed by *in vitro* binding site selection experiments (Blackwell and Weintraub, 1990; Mavrothalassitis *et al.*, 1990; Pollock and Treisman, 1990; Verrijzer *et al.*, 1992; Kroeger and Morimoto, 1994). We have taken this approach one step further, to define the optimal binding site of the CytR repressor, alone as well as in the presence of its helper protein, cAMP–CRP. We find that CytR by itself preferentially binds two octamer repeats, in direct or inverted orientation, and separated by 2 bp. Surprisingly, these DNA configurations are not optimal for co-binding with cAMP–CRP. Rather, cAMP–CRP stabilizes alternative DNA-binding modes of CytR.

In one mode, CytR recognizes inverted octamer repeats separated by 10–13 bp; in another the repressor binds direct octamer repeats separated by 1 bp. Thus, CytR is a very adaptable DNA-binding protein that retains a high degree of flexibility even in the presence of its co-repressor.

Results

Selection strategies

To define the optimal binding site for the CytR repressor, we used a PCR-based binding site selection assay (see Materials and methods). The starting material for the experiments was a population of ~10^{11} DNA fragments carrying *deoP2* promoter sequences in which the central 27 bp, containing the CytR-binding site, had been randomized. In a first set of experiments, the pool of DNA fragments was incubated with CytR, and an electrophoretic mobility shift assay was performed. DNA was isolated from the band containing CytR–DNA complexes and amplified by error-prone PCR, in order to increase the DNA sequence diversity before the next round of selection. After several rounds, the isolated DNA fragments were cloned, sequenced and characterized by mobility shift and footprinting assays. To investigate whether cAMP–CRP has any effect on the DNA recognition specificity of CytR, we repeated the selection/amplification experiments in the presence of both CytR and cAMP–CRP.

Isolation of the optimal CytR-binding site

The course of the selection for CytR operators (in the absence of cAMP–CRP) was evaluated by mobility shift assays (data not shown). The analyses revealed that after three rounds of selection, the enriched pool of DNA sequences has a higher affinity for CytR than *deoP2* wild-type (wt) fragments, and that near-optimal binding sites for the CytR repressor are obtained after 4–5 rounds of selection. Hence, fragments obtained after 6–8 and 12 rounds were cloned and sequenced (Figure 1). All the sequences show homology to the octamer motifs 5′-AATG$^{T}/_{C}$AAC-3′ and 5′-GTTGCATT-3′. We have termed these left (L) and right (R) half-sites, respectively. Based on the orientation and number of octamer repeats, the selected fragments were divided into four groups (Figure 1).

Of 46 recovered fragments, two were identical. The majority (33 fragments) contain two inverted octamer repeats, separated by 2 bp (Group A1). Additionally, seven sequences contain inverted repeats separated by 1 bp; most of these isolates were obtained after a few (6–8) rounds of selection. Three sequences contain a direct repeat arrangement of two octamer motifs (Group A2), and one sequence contains three repeats (Group A3). Finally, two sequences (obtained after six or seven rounds of selection) carry only one repeat (Group A4).

The consensus sequence of group A1 is a near perfect palindrome, 5′-AATG$^{T}/_{C}$AAC-GC-GTTGCATT-3′. There seems to be a preference for asymmetry. First, only one entirely symmetric 18 bp sequence (AATGTAAC-GC-GTTACATT) was obtained, while the consensus sequence (AATGTAAC-GC-GTTGCATT) was found in eight copies. Second, the L half-sites are more diverse than the R half-sites among A1 sequences. Finally, the octamer boxes in most of the fragments reside in the right end of the randomized 27 bp region. Presumably, the two thymidines at the 3′ end of the recognition motif are provided by the constant DNA region next to the randomized portion (see Figure 1). This may have biased the selection in favor of R versus L direct repeats.

The affinity of CytR for individual DNA sequences was determined by mobility shift assays (Figure 3). Isolates obtained after 12 rounds of selection bind with almost indistinguishable affinities to CytR, and have 30-fold higher affinity for the repressor than the *deoP2*-wt promoter. The sequence A08-21 (L–1–R) with a spacing of 1 bp between the octamer motifs has a 2-fold lower affinity for CytR than the L–2–R or R–2–R sequences; DNA fragments containing a single octamer repeat bind CytR half as efficiently as the *deoP2*-wt promoter. Finally, A12-29 has an affinity for CytR similar to that of other sequences obtained after 12 selection rounds, despite the presence of an extra octamer repeat in this sequence.

To gain insight into the qualitative interactions of CytR with DNA, we performed DNase I and dimethyl sulfate (DMS) footprinting at repressor concentrations that saturate the binding site. The DNase I footprints serve to delineate the binding sites; DMS footprinting identifies purines in close contact with protein. The salient features of the footprints of sequences A12-09 and A12-44, representing the inverted (L–2–R) and direct (R–2–R) repeat configuration, are as follows (Figure 4). The DNase I footprint covers ~20 bp at both sequences, and, as expected, CytR interacts specifically with both operator half-sites. Interaction of CytR with the octamer motifs AATGTAAC and GTTGCATT invariably protects the central guanine from DMS methylation, consistent with the well conserved G at this position (Figure 1). Methylation of the adenine at position 6 is only observed in the L half-site of A12-09.

Isolation of DNA sequences that support CRP_2–CytR–DNA complex formation

We next performed the selection in the presence of both CytR and cAMP–CRP. As the two protein species bind cooperatively to *deoP2*, even when the entire sequence between the two CRP targets has been randomized (Søgaard-Andersen and Valentin-Hansen, 1993), we were able to obtain a discrete band containing the quaternary CRP_2–CytR–DNA complex (Figure 2), in the first round of selection. Analytical mobility shift assays showed that after eight rounds of selection the DNA pool forms the CRP_2–CytR–DNA complex more efficiently than the *deoP2*-wt promoter (data not shown).

Sequencing revealed 34 unique DNA fragments out of a total of 36 recovered. Many different CytR-binding site configurations seem to allow cooperative interaction of CytR with cAMP–CRP. We have divided them into three groups (Figure 2). The largest group, consisting of 23 sequences (Group C1), contains near perfect palindromes with the consensus 5′-$^{AC}/_{TG}$TGCAAC-N_x-GTT$^{G}/_{A}$CATT-3′, x = 10, 11, 12 or 13. The 5′ end of the motif is either AC or TG, suggesting that these nucleotide steps share a common feature of importance for repression complex formation at C1 sequences. Six sequences have two octamer boxes separated by 1 bp, in either R–1–R or L–1–L direct repeat configuration (Group C2). The

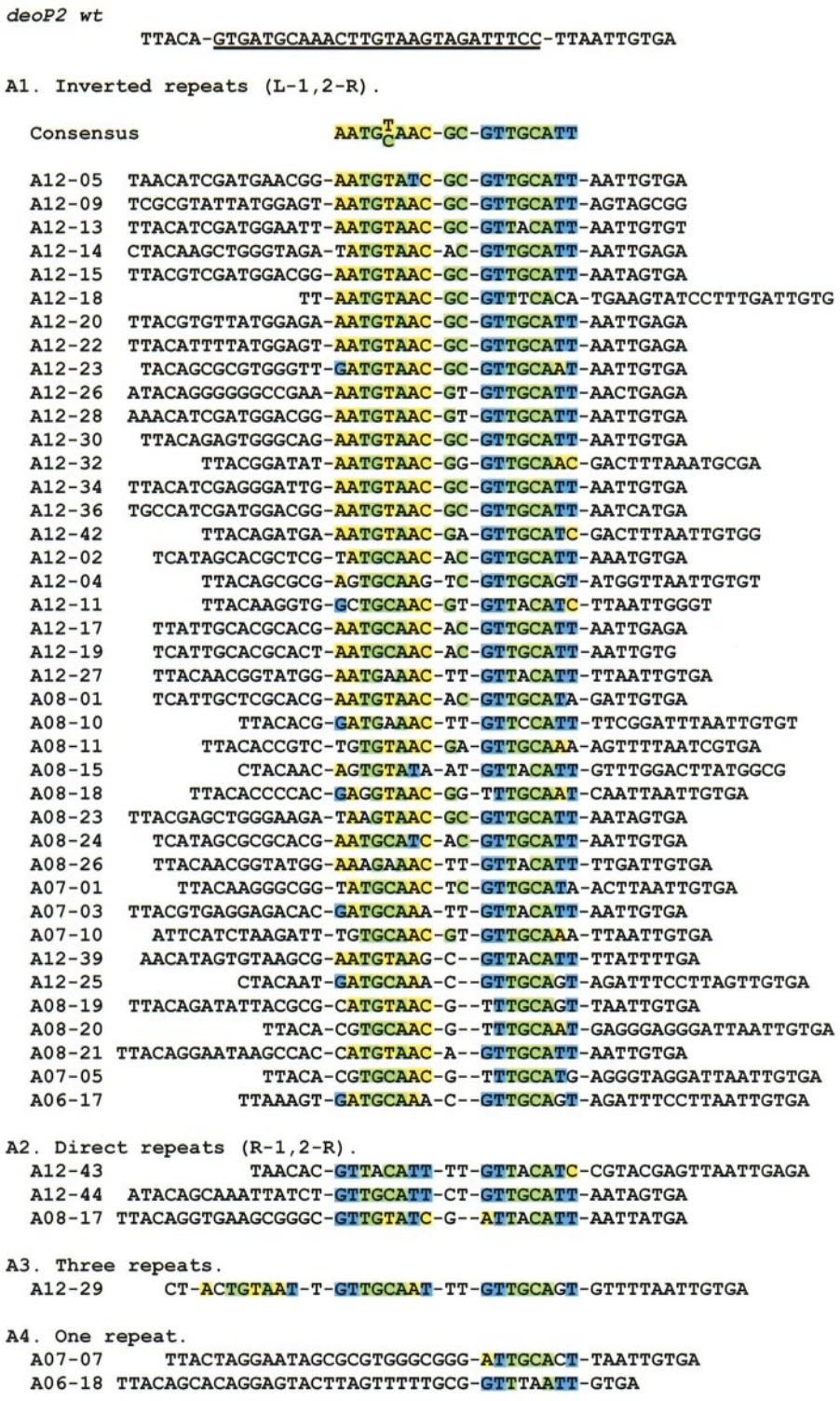

Fig. 1. Isolation of CytR-binding sites. The *deoP2*-wt sequence is shown at the top; the originally randomized region is underlined. Sequences obtained from one experiment were divided into four groups, and gaps (–) introduced to display homologies. Bases are colored according to the consensus of Group A1: green color indicates the four bases that are identical in both halves of the consensus (--TGCA--); yellow and blue indicate bases specific for left (AA--T-AC) and right (GT---TT) half-operators, respectively. Numbers to the left of the sequences indicate the number of selection rounds performed (e.g. '12'), followed by a serial number (e.g. '05'). 'A' indicates that the selection was performed in the absence of cAMP–CRP. PCR amplification primers (deoprim1 and deoprim2) anneal immediately upstream and downstream of the presented sequences, respectively. Two copies of the A12-14 sequence was recovered; otherwise, all sequences were different.

thymidine at the 3′ end of R half-sites seems to be provided by the constant region next to the originally randomized region, and may have biased the selection in favor of C1 sequences. No everted (R–L) sequences were isolated. Finally, five DNA isolates contain three repeats, in various orientations (Group C3). The two outermost octamer motifs of this group are inverted repeats separated by 10–14 bp; moreover, three of the C3 sequences bear

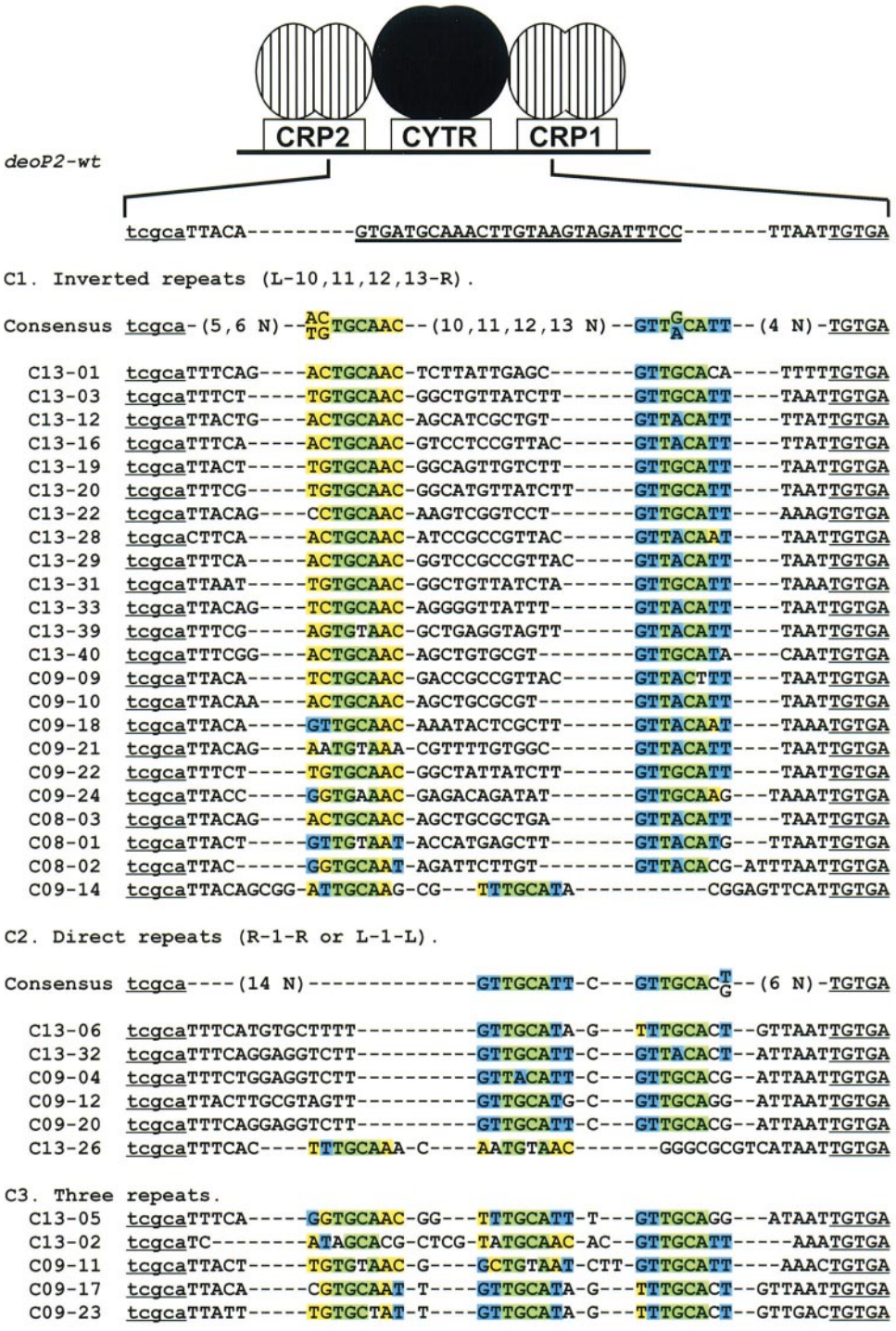

Fig. 2. Isolation of CytR-binding sites that support CRP_2–CytR–DNA complex formation. The *deoP2*-wt sequence is shown at the top. The CRP2- and CRP1-binding sites are centered at –93.5 and –40.5, respectively, relative to the transcription initiation site. cAMP–CRP complexes are hatched; CytR is in black. The CRP_2–CytR–DNA repression complex covers ~80 bp on one face of the DNA helix, and is held together by CRP–DNA, CytR–DNA and CRP–CytR interactions (Pedersen *et al.*, 1991; Søgaard-Andersen *et al.*, 1991a,b; Rasmussen *et al.*, 1993). The originally randomized portion is indicated by thick underlining; the flanking half-site recognition motifs of the CRP targets are indicated by thin underlining (the centers of CRP2 and CRP1 are 3 bp upstream and downstream, respectively, of the sequences shown). Sequences from one selection experiment were divided into three groups, and gaps (–) introduced to show homologies. Bases are colored according to the consensus of Group C1: green indicates consensus bases that are identical in the two half-sites (–TGCA–); yellow and blue indicate bases specific for left (AC----AC or TG----AC) and right (GT-A--TT) half-operators, respectively. The PCR amplification primers (deoprim1 and deoprim2) anneal upstream of and including the tcgca sequence, and immediately downstream from the shown sequence, respectively. Numbers to the left of the sequence indicate the number of selection rounds (e.g. '13') followed by a serial number (e.g. '01'). 'C' indicates that this selection was done in the presence of cAMP–CRP. Two copies each of C13-32 and C13-03 were recovered; the remaining sequences are different.

two octamer boxes in a direct repeat arrangement, and separated by 1 bp. Thus, group C3 sequences show homology to both C1 and C2 sequences.

There seem to be rather strict rules for the position of the L and R CytR half-sites relative to the flanking CRP sites. For C1 sequences, the separation between the TCGCA motif of CRP2 and the L repeat is almost exclusively 5 or 6 bp, and the spacing between the R repeat and TGTGA of CRP1 is 4 bp. Correspondingly, for C2 sequences, the R–1–R and L–1–L repeats are separated from the CRP2 and CRP1 targets by 14/6 and 6/14 bp, respectively. Thus, the C13-26 sequence with the L–1–L arrangement is basically the inversion of the R–1–R sequences.

The independent affinity of CytR for the C-sequences is considerably lower than for the A-sequences (2- to 5-fold lower; Figure 3, black bars). Independent binding of CytR to the selected sequences thus produces a hierarchy of affinities, A-operators>C-operators>*deoP2*-wt-operator. In the presence of cAMP–CRP, CytR binds the C1, C2 and C3 sequences with very similar affinity (Figure 3, hatched bars). However, the cooperativity exhibited by the C-fragments (~20- to 50-fold) is less than that of the *deoP2*-wt fragment (~160-fold). As a result, the cAMP–CRP-dependent affinity of CytR for the C-sequences is only 2-fold higher than for *deoP2-wt*. Thus, the high independent affinity of CytR for the C-sequences, relative to *deoP2*-wt, has not resulted in a corresponding increase in cAMP–CRP-dependent affinity.

We could not find A-sequences for which it would make sense to test for cooperative binding, since most of the A-fragments contain mutations in the CRP targets, or because the binding sites are too close to allow simultaneous binding of cAMP–CRP and CytR. However, C09-14 from the cAMP–CRP-dependent selection (Figure 2, last sequence in Group C1) contains two inverted repeats separated by 2 bp, like sequences from the independent selection (Group A1). The cooperative binding of CytR and cAMP–CRP to this sequence is very inefficient, even though the independent CytR affinity is similar to that of other C-sequences (Figure 3). This implies that it is the configuration of the CytR half-sites and the position of these half-sites relative to the flanking CRP targets, and not the strength of the CytR–DNA interaction, that are important for cooperative binding of CytR and cAMP–CRP.

Two sequences from each of groups C1, C2 and C3 were footprinted by DNase I and DMS in the presence of CytR and cAMP–CRP (Figure 4). As observed for the A-sequences, CytR protects the central guanine of the octamer motifs from DMS methylation, and only L half-sites exhibit increased DMS reactivity of the adenine at position 6. Notably, the footprint patterns in the CytR region are relatively independent of the presence of cAMP–CRP. For example, in the DNase I footprints, the C1 sequences (C13-20 and C13-40) exhibit a 3 bp unprotected region in the middle of the CytR operator, and C09-20 (R–1–R) and C13-26 (L–1–L) contain unprotected regions to the left or right of the operator, regardless of the presence of cAMP–CRP. These results imply that a CytR protein bound to C-sequences does not change shape upon addition of cAMP–CRP. It thus appears that selection in the presence of cAMP–CRP has produced DNA targets

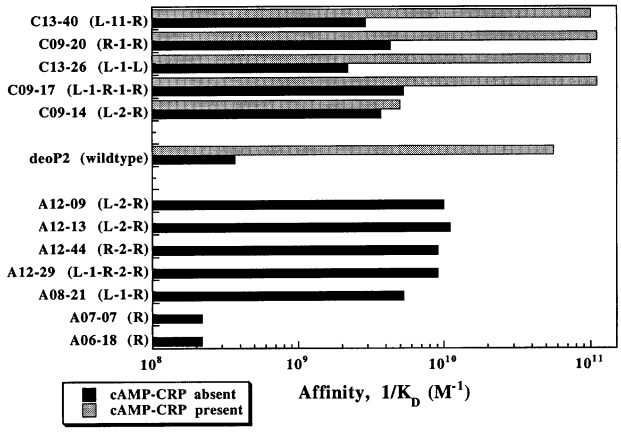

Fig. 3. Apparent CytR affinity ($1/K_D$) for the selected DNA. The apparent affinity of CytR for selected DNA sequences and *deoP2*-wt, in the absence (black bars) or presence (gray bars) of 6×10^{-9} M CRP, was determined by the mobility shift assay (see Materials and methods). The apparent dissociation constant, K_D, was taken as the CytR concentration that binds 50% of the DNA fragments. CytR was in >100-fold excess to the binding site (cAMP–CRP absent); in the presence of cAMP–CRP, both proteins were in at least 5-fold excess to the binding sites. The cAMP concentration was 50 μM whenever CRP was employed. Note that previously published protein–DNA affinities were determined by footprinting; the mobility shift assay gives considerably higher affinities for the interaction of CytR with DNA (Pedersen *et al.*, 1991, 1992, 1995).

to which CytR binds in a conformation that is designated to interaction with cAMP–CRP.

Independent binding of CytR to isolates containing three octamer boxes does not result in the simple pattern described above. The DNase I footprints are extended, and cover all three repeats; correspondingly, all three repeats exhibit DMS protection. This is presumably caused by simultaneous binding of two CytR molecules, or by a mixed population of complexes in which one CytR binds either two of the three repeats. The combined footprints on these isolates, however, resemble those of group C1 and C2 sequences: Addition of cAMP–CRP to the C09-17 sequence creates a C09-20 (R–1–R)-like footprint, and the C13-05 footprint resembles those of C13-40 (L-11-R) and C13-20 (L-13-R).

Discussion

Previous studies have revealed that protein–DNA and protein–protein interactions, as well as protein-induced DNA-bending, cooperate in an organized manner to form repression complexes at CytR-regulated promoters (Søgaard-Andersen *et al.*, 1991a; Pedersen *et al.*, 1992; Søgaard-Andersen and Valentin-Hansen, 1993). In these complexes, the DNA bends strongly around cAMP–CRP, bringing the DNA-bound regulators into close proximity (Søgaard-Andersen *et al.*, 1991b; Crothers and Steitz, 1992). We show here that besides acting as an architectural element, and providing contacts through its surface, CRP also alters the DNA-binding mode of the CytR repressor. Furthermore, our results reveal that widely different conformations of the CytR repressor can cooperate with cAMP–CRP to form nucleoprotein complexes of equal stability. Below we discuss these novel aspects and their implications for CRP/CytR combinatorial regulation and multiprotein–DNA complex formation in general.

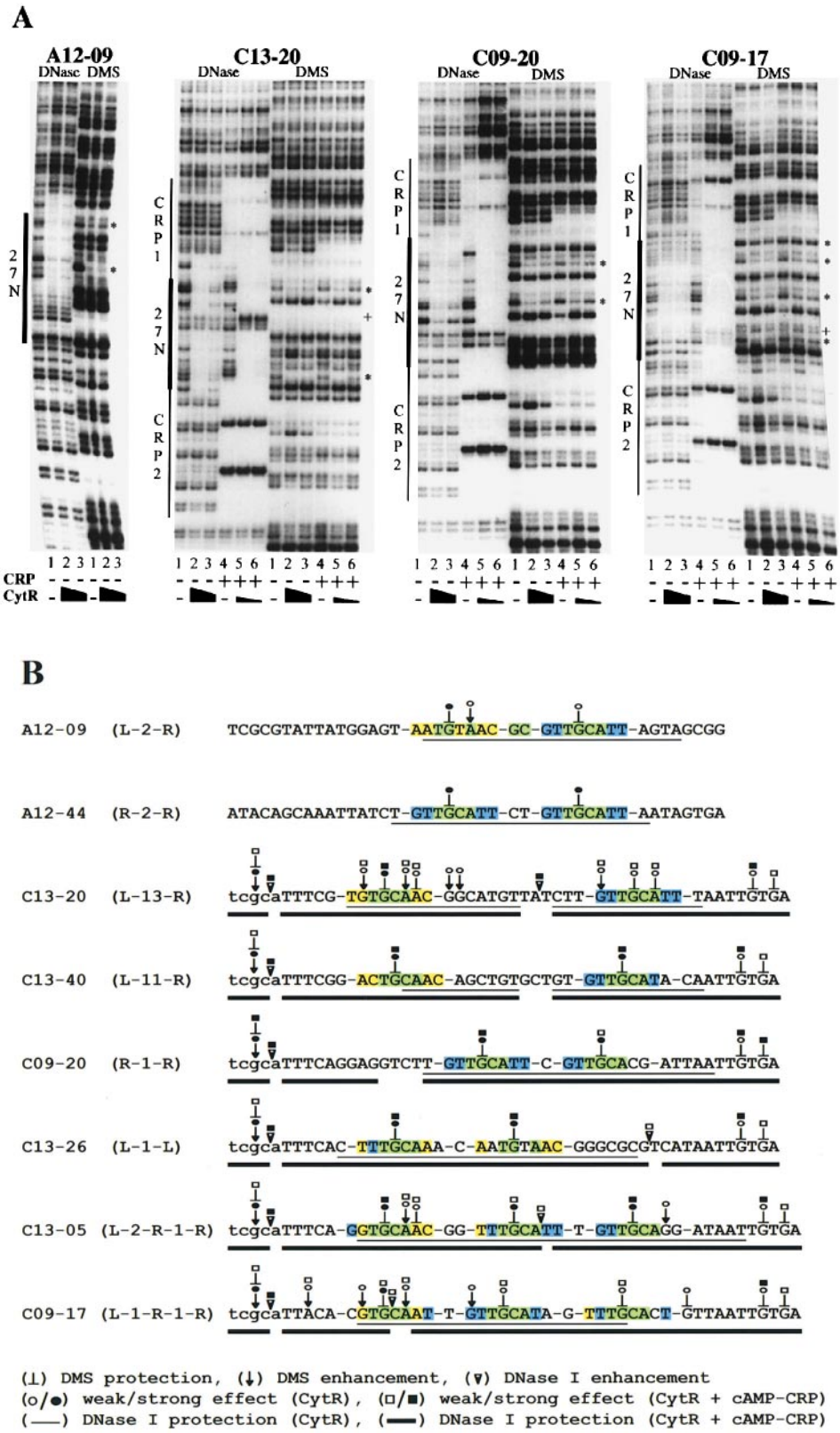

Fig. 4. DNase I and DMS footprints of CytR and cAMP–CRP. (**A**) Footprinting of A12-09, C13-20, C09-20 and C09-17. The left and right halves of the panels show DNase I and DMS footprints, respectively. (*) denotes guanines protected from DMS methylation by CytR; DNase I enhancements in the CytR-binding region are indicated by (+). Final CRP concentrations: lanes 1–3, no CRP added; lanes 4–6, 5×10^{-8} M CRP. Final CytR concentrations: lanes 1, no CytR; lanes 2, 5×10^{-8} M CytR; lanes 3, 10^{-8} M CytR; lanes 4, no CytR; lanes 5, 5×10^{-9} M CytR; lanes 6, 5×10^{-10} M CytR. cAMP was added to 50 μM in reactions containing CRP. The initially randomized region (27 bp), and the CRP1 and CRP2 targets are indicated to the left of each panel. (**B**) Schematic representation of DNase I and DMS footprints of sequences isolated in the absence of cAMP–CRP (A12-09 and A12-44) or in its presence (C13-20, C13-40, C09-20, C13-26, C13-05 and C09-17). Circles and squares indicate positions affected by independent CytR binding and by formation of the CRP_2–CytR–DNA complex, respectively. Empty/filled symbols indicate weak/strong effects. Nucleotides protected from DNase I cleavage by CytR, or by formation of the CRP_2–CytR–DNA complex, are shown by thin and thick underlining, respectively. Coloring is as in Figure 1 (A-sequences) or Figure 2 (C-sequences).

cAMP–CRP changes the DNA-binding mode of CytR

The present work allows us to define an 8 bp half-operator consensus (ATTG$^{T}/_{C}$AAC) for CytR. In the absence of cAMP–CRP, the CytR repressor binds two such octamer motifs in inverted or direct repeat arrangement, preferably separated by 2 bp (Figure 6A). The CytR half-operator consensus for cooperative binding with cAMP–CRP is slightly different. The differences are at the left edges of the CytR operator, and might facilitate wrapping of the DNA helix around the proteins in the combined complex. Alternatively, since the strong CRP2 target is expected to span ~30 bp (Liu-Johnson *et al.*, 1986), the sequence at the edges of the CytR operator may optimize both cAMP–CRP and CytR interactions with DNA. Nevertheless, the absence of major changes to the half-operator consensus indicate that cAMP–CRP only minimally interferes with the structures of the individual DNA-binding domains of CytR. cAMP–CRP does, however, induce drastic changes in CytR's quaternary structure. Thus, cAMP–CRP preferentially stabilizes a set of CytR conformations that fit operators composed of inverted repeats with wide spacing (10–13 bp), or direct repeats separated by 1 bp. As illustrated in Figure 6B and C, these DNA arrangements are expected to bind either a roughly symmetrical conformation of the CytR dimer in which the DNA-binding domains are held in an inverted orientation with their recognition centers approximately two DNA helical turns apart, or an asymmetric conformation in which the DNA-binding motifs are in direct orientation, with their centers approximately one helical turn apart.

Generality of the isolated DNA sequences

One complication in devising a general selection scheme for cAMP–CRP-dependent CytR-binding sites was that cooperative binding of cAMP–CRP and CytR had been observed on DNA templates with 52–54 bp separations between the centers of the two CRP targets (Søgaard-Andersen *et al.*, 1990; Pedersen *et al.*, 1991; Holst *et al.*, 1992). Thus, variation in rotational and translational separation of the flanking CRP sites could potentially affect CytR's DNA-binding mode, and, consequently, conclusions derived in the context of one promoter might not necessarily apply to others. This potential problem was circumvented by employing error-prone PCR during each selection/amplification round. In good agreement with results obtained for natural promoters, the selected C-sequences have CRP–CRP distances of 51–54 bp (Figure 2). Therefore, it is plausible that the conclusions established from the biochemical experiments with *deoP2* apply *in vivo* as well, and should be applicable to other promoters of the CytR regulon.

In fact, natural CytR-binding sites exhibit a striking homology to the cAMP–CRP-dependent binding sites (C-sequences) identified in the present work (Figure 5). The CytR operator of the *cdd* promoter contains R–1–R direct repeats, and the repeats are separated by 5 and 14 bp, respectively, from the flanking CRP targets. This is very similar to the arrangement of group C2 sequences (Figure 2). Also, the CytR-binding site in the *cytR* promoter (L–1–L) matches the C2 sequences. *nupG* is a representative of group C1 sequences, and the arrangement of the three CytR repeats in the divergent *cytX-rot* promoter region corresponds to the group C3 sequences. Finally, the CytR operator site of the wild-type *deoP2* promoter deviates somewhat from the consensus sequences. Thus, the CytR half-operators in *deoP2* are immediately adjacent, and their orientation is not obvious. However, the relative positions of the CytR- and CRP-binding sites resemble those of group C2 sequences. Taken together, the configurations of the natural CytR operators and the selected C-sequences are very similar. Thus, the heterogeneous nature of natural CytR operators, which appeared at first rather puzzling, can now be understood in detail. In this regard, we note that combination of many sub-optimal interactions seems to be inherent in the design of gene regulatory systems composed of multiple factors. The CRP/CytR regulatory system obviously contains many adjustable parameters; for a given natural promoter, only a subset of these is optimized, thus preserving a dynamic regulatory circuit.

Fig. 5. Natural CytR-binding sites. The sequences of five CytR-regulated promoters are compared with the cAMP–CRP-dependent CytR consensus determined in this report. Coloring is as in Figure 2. One half-site recognition sequence in each of CRP2 and CRP1 is underlined; the CRP consensus sequence is TGTGA-N_6-TCACA (de Crombrugghe *et al.*, 1984; Ebright *et al.*, 1989). The centers of the flanking CRP targets are 3 bp upstream and downstream from the shown sequence. CRP- and CytR-binding sites were determined by *deoP2* (Valentin-Hansen, 1982; Pedersen *et al.*, 1991; Rasmussen *et al.*, 1993), *cdd* (Holst *et al.*, 1992), *cytR* (Pedersen *et al.*, 1992); *nupG* (Pedersen *et al.*, 1995); and *cytX-rot* (Nørregaard-Madsen *et al.*, 1994).

Structural implications: the CytR repressor protein

CytR is a dimer in solution and when bound to the *deoP2* promoter (H-H.Kristensen *et al.*, 1996). On the basis of the characteristics of the selected operators, the footprinting patterns and the migration rate of the CytR–DNA and CRP_2–CytR–DNA complexes in the gel mobility shift experiments (this study; data not shown), we conclude that CytR also binds as a dimer to the selected A- and C-sequences. CytR belongs to the LacI repressor family and exhibits extensive amino acid sequence homology to several other members (e.g. PurR, LacI and GalR; Weickert and Adhya, 1992). Based on the PurR–DNA co-crystal structure (Schumacher *et al.*, 1994), and three crystal structures of LacI (as free protein and in complex with operator DNA or inducer; Lewis *et al.*, 1996), CytR is expected to consist of an N-terminal DNA-binding domain of ~60 amino acids, connected via a hinge region to a ~270 amino acid C-terminal domain that mediates dimerization and ligand binding.

The ability of CytR to contact operators with widely different half-site spacings reveals a rotational and translational flexibility that is unprecedented in the LacI family.

What structural information allows CytR to interact with DNA in such a relaxed manner? Several lines of evidence indicate that the flexibility resides solely in the hinge that links the DNA-binding domain to the ligand-binding/dimerization core. The amino acids involved in dimerization of CytR have been determined genetically (Barbier and Short, 1993), and correspond well with amino acids involved in dimerization of other LacI members (Schumacher *et al.*, 1994; Lewis *et al.*, 1996; Friedman *et al.*, 1995; for review, see Weickert and Adhya, 1992). It thus seems unlikely that CytR should possess alternative dimerization interfaces between which the subunits could slide to allow the DNA-binding domains to recognize differently spaced half-sites. Furthermore, the high sequence homology among the HTH motifs of LacI members (60% amino acid identity among PurR and CytR HTHs), as well as between their DNA operator half-sites (Schumacher *et al.*, 1994), strongly suggests that the HTH motifs take very similar structures, and that they interact with their half-operators in very similar ways. Thus, there appears to be no special feature in the DNA-binding domain of CytR that could account for CytR's DNA-binding flexibility.

We thus favor a model in which the interdomain linker of CytR provides the observed flexibility. The following observations support this idea. Proteolysis and NMR studies revealed that the interdomain linker of two of CytR's closest relatives, the PurR and LacI repressors, is disordered in the absence of operator DNA (Wade-Jardetzky *et al.*, 1979; Choi and Zalkin, 1994; Nagadoi *et al.*, 1995). Upon interaction with specific operator DNA, this region forms a stable DNA-binding unit consisting of two α-helical hinges, one from each monomer (Schumacher *et al.*, 1994; Lewis *et al.*, 1996). Thus, PurR (and LacI) contains two different DNA-binding motifs. One is the HTH motif, which fits snugly into the major groove of one half-operator; the other is composed of the two hinge regions which form a structural lever that prises open the DNA minor groove at the center of the operator (Schumacher *et al.*, 1994). This results in a 45° kink of the DNA away from the protein. Two key features of the PurR–DNA complex make this possible. The Leu54 residues of each monomer hold the hinge helices together, so that their side chains can intercalate into the DNA; the CpG step at the center of the operator facilitates bending of the DNA towards the major groove (Gartenberg and Crothers, 1988), and provides minor groove contacts for the Leu54 side chains. Clearly, CytR cannot bind the C-sequences (or the natural targets) in a similar binding mode, since the hinge helices would be pulled too far apart when the HTH motifs interact with direct repeats, or inverted repeats separated by one DNA helical turn (Figure 6B and C). In support of this notion, all LacI members, but not CytR, carry a leucine at the position corresponding to Leu54 of the PurR repressor (Weickert and Adhya, 1992). Moreover, inspection of the recognition sequences reveals the conservation of the central CpG dinucleotide in all LacI family operators (Schumacher *et al.*, 1994), again with the exception of CytR operators. It seems likely, therefore, that the interdomain linker serves a different function in CytR. Conceivably, the hinge regions of the DNA-bound CytR dimer might be structurally disordered and only loosely or not at all tethered to one another. This would afford the DNA-binding domains of CytR with considerable freedom of movement, and allow them to interact with half-operators separated by as little as 1 bp or as much as 13 bp, contact half-operators on opposite faces of the DNA helix, or flip 180° in order to interact with repeats in either orientation (Figure 6).

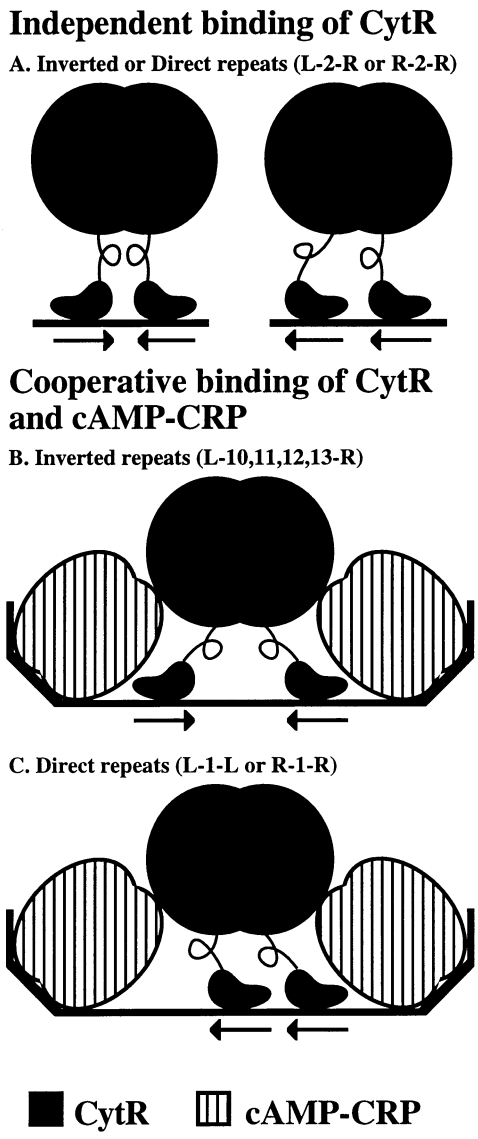

Fig. 6. cAMP–CRP-induced conformational changes of the CytR repressor. (**A**) In the absence of cAMP–CRP, the CytR repressor preferentially binds octamer repeats in direct or inverted orientation, separated by 2 bp. In the presence of cAMP–CRP, CytR recognizes (**B**) inverted octamer repeats separated by 10–13 bp or (**C**) direct repeats separated by 1 bp.

We can now consider why CytR requires cAMP–CRP to bind DNA efficiently. As the present study shows, it is possible to increase DNA binding of the repressor considerably (~30-fold), simply by changing the operator sequence. However, the affinity of CytR for its optimal operator is still considerably lower than the affinities of most prokaryotic repressors for their natural operator sequences. Moreover, the affinity of CytR for a DNA fragment carrying two repeats in consensus arrangement (L–2–R) is only 50-fold higher than for a fragment harboring one repeat. Thus, the cooperative binding of

the two DNA-binding domains of CytR is low. Presumably, fixation of the floppy DNA-binding domains by binding to the operator is entropically costly and, hence, lowers the affinity relative to a hypothetical rigid protein–DNA interaction. Thus, a reduced intrinsic affinity for its operator may be the price CytR has to pay for increased structural flexibility.

Structural flexibility and induced fit

While it is not surprising that transcription factors can exhibit promiscuous DNA-binding characteristics, or undergo induced fit when interacting with other molecules, it is remarkable that CytR can assume multiple conformations of about equal stability when co-binding with cAMP–CRP (Figure 6). This scenario, however, leaves some difficult questions unanswered. First, why are the A-sites, which exhibit the highest affinity for CytR, not used for cooperative complex formation? A few simple possibilities are likely: binding of CytR to A-sites may deform the DNA in a way that impedes wrapping the DNA helix around the proteins in the CRP_2–CytR–DNA complex, or it may induce DNA conformational changes in the flanking CRP targets, unfavorable for binding of cAMP–CRP. Alternatively, CytR may adopt a conformation at these sites that is incompatible with positive CRP–CytR interactions.

Second, what is the mechanism of the cAMP–CRP-induced fit? A clue to this issue may lie in the arrangements of the binding sites. Thus, every C-arrangement contains at least one CytR half-operator in proximity to a CRP target, and always in the same orientation. This suggests that there are contacts between the DNA-binding domain of CytR (the 'toes' of CytR; Figure 6B and C) and CRP near the DNA, besides the interaction between the ligand-binding/dimerization domain of CytR and CRP demonstrated previously (Søgaard-Andersen and Valentin-Hansen, 1993). Alternatively, CRP-induced distortion of the DNA helix at the edges of the CRP-binding sites may present an optimal target for a CytR DNA-binding domain. Finally, it is possible that the contact between cAMP–CRP and the ligand-binding/dimerization domain of CytR triggers a conformational change through CytR that shifts it from a state with preference for A-sequences to one with a preference for C-sequences. We cannot distinguish at present between these models.

General implications

Flexible interdomain linkers play a key structural and functional role in a number of DNA-binding proteins (e.g. AraC, Eustance *et al.*, 1994; α2 yeast repressor, Smith and Johnson, 1992; POU domain proteins, reviewed by Herr and Cleary, 1995; COUP-TFs, Cooney *et al.*, 1992; α subunit of *Escherichia coli* RNA polymerase, Blatter *et al.*, 1994), and may be a rather general feature of transcription factors. Should we also expect CytR's DNA-binding promiscuity in multiprotein–DNA complexes to be a general phenomenon? For proteins that rely on multiple protein–protein and protein–DNA interactions to target regulatory regions of different binding site composition, the answer is probably yes. Formation of complex nucleoprotein structures involves the alignment in three-dimensional space of the individual proteins with their DNA targets, and of the epitopes involved in protein–protein interactions, as well as the deformation of the DNA helix. The seemingly very difficult task of aligning all these molecules so that their recognition elements can interact productively should be greatly facilitated by elasticity in one or more of the partners involved, as this would help to minimize the energetic cost of the conformational transitions needed. Also, the ability of a DNA-binding protein to adopt different conformations, dependent on the configuration of the binding site, could provide selectivity in association with regulatory cofactors. Thus, an adaptable DNA binder would be an attractive target for different co-regulators, and provides a simple device for constructing versatile combinatorial systems.

Materials and methods

Mobility shift assay

^{32}P-Labeled DNA fragments and proteins were incubated in binding buffer [10 mM Tris–HCl (pH 7.8), 50 mM KCl, 1 mM EDTA, 50 μg/ml acetylated bovine serum albumin, 1 mM dithiothreitol, 0.05% NP-40, 50 μM cAMP, 15 μg/ml non-specific competitor DNA (pGEM4, obtained from Promega)] for 30 min (Figure 3), or as described below (selection assays), in a total volume of 10 μl (Pedersen *et al.*, 1991). Three μl of loading buffer (binding buffer containing 50% glycerol and 0.1 mg/ml bromophenol blue) was added, and the samples immediately loaded on the gel with current on; electrophoresis was at 200 V for 75 min. Gels were 5% polyacrylamide prepared from a 44:0.8% (acrylamide: *N,N′*-methylenbisacrylamide) stock. The electrophoresis buffer employed was 10 mM Tris–HCl (pH 7.8), 1 mM EDTA, 10 μM cAMP. Following electrophoresis, the gel was dried and autoradiographed.

In vitro binding site selection

We modified the binding site selection assays described previously (Blackwell and Weintraub, 1990; Mavrothalassitis *et al.*, 1990; Pollock and Treisman, 1990). Double-stranded template, obtained by annealing ~2.5 pmol of the oligo 'deo27N′ to excess ^{32}P-labeled oligo 'deoprim2' and filling out with Klenow enzyme, was incubated at 37°C with protein(s) in a total volume of 10 μl, and subjected to a mobility shift assay (see above). In each round of selection, the protein concentrations were adjusted so that a little, but significant amount of the relevant protein–DNA complex (CytR–DNA or CRP_2–CytR–DNA complex) would form. For the independent binding of CytR, this corresponded to 7×10^{-8} M and 7×10^{-10} M CytR in the first and final selection round, respectively. For the cooperative binding with cAMP–CRP, the corresponding CytR concentrations were 7×10^{-10} M and 2×10^{-11} M. CRP and cAMP were employed at concentrations of 2×10^{-7} M and 50 μM, respectively. In the initial rounds of selection, incubation proceeded for 2–3 h; in later rounds, incubation was extended to 48 h. The extended incubations permitted the proteins to go through many association and dissociation events, leading to isolation of high-affinity DNA sequences. Following electrophoresis, the band corresponding to the protein–DNA complex of interest (CytR–DNA or CRP_2–CytR–DNA) was cut out of the dried gel, and the DNA eluted from the gel slice in 0.5 M NH_4Ac, 1 mM EDTA, 0.1% SDS, 10 mM Tris–HCl (pH 7.3), 10 mM $MgCl_2$, overnight at 37°C. After phenol/chloroform extraction and ethanol precipitation, the DNA was amplified by error-prone PCR (for another round of selection) or standard PCR (for cloning), using the primers deoprim1 and deoprim2. Standard and error-prone PCR were performed according to Cadwell and Joyce (1992), except that primer concentrations were adjusted so that <10% was incorporated during the PCR. Approximately 20 PCR cycles were performed; PCR products were acrylamide gel-purified before further use.

DNase I and DMS footprinting

DNase I experiments were carried out as described by Galas and Schmitz (1978), with the modifications described in Pedersen *et al.* (1991). DMS experiments were performed according to Vidal-Ingigliardi *et al.* (1991), with the modifications of Pedersen *et al.* (1995).

DNA oligos

The primers used were: deoprim1, 5′-CCGAATTCCCTTTGAAAGTG-AATTATTTGAACCAGATCGCA-3′; deoprim2, 5′-GAGATCCGG-AACACACTTCGATACACA-3′; deo27N, 5′-AGTGAATTATTT-GAACCAGATCGCATTACANNNNNNNNNNNNNNNNNNNNN-

NNNNTTAATTGTGATGTGTATCGAAGTGTGTTCCGGAT-3′. N denotes any base. The *Eco*RI and *Bsp*E1 restriction sites are underlined.

Plasmids and preparation of ^{32}P-labeled DNA fragments

p13-134(8) is a derivative of p13-134 (Søgaard-Andersen *et al.*, 1990), and contains wild-type *deoP2* sequence from –116 to +98, except for a mutation at position –25 (G→C) which creates a *Bsp*E1 restriction site. DNA isolates recovered in the selection assay were PCR amplified with deoprim1 and deoprim2, digested with *Eco*RI and *Bsp*E1, and cloned into *Eco*RI–*Bsp*E1-restricted p13-134(8). The 300 bp DNA fragments used in the footprinting (Figure 4) and mobility shift experiments (Figure 3) were prepared by PCR with pUC reverse sequencing primer (–24) and ^{32}P-labeled pUC sequencing primer (–40) (New England Biolabs), using p13-134(8) derivatives (carrying the recovered DNA sequences from the selection experiments) or p13-134 (encoding wild-type *deoP2*) as templates.

Proteins

CytR and CRP proteins were purified according to Pedersen *et al.* (1991) and Ghosaini *et al.* (1988), respectively.

Acknowledgements

We thank Eva Welander Raabjerg for enthusiastic and expert technical assistance. This work was supported by the Danish Center for Microbiology, and the Center for Interaction, Structure, Function, and Engineering of Macromolecules.

References

Barbier,C.S. and Short,S.A. (1993) Characterization of *cytR* mutations that influence oligomerization of mutant repressor subunits. *J. Bacteriol.*, **175**, 4625–4630.

Blackwell,T.K. and Weintraub,H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science*, **250**, 1104–1110.

Blatter,E.E., Ross,W., Tang,H., Gourse,R.L. and Ebright,R.H. (1994) Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell*, **78**, 889–896.

Cadwell,R.C. and Joyce,G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Applications*, **2**, 28–32.

Choi,K.Y. and Zalkin,H. (1994) Role of the purine repressor hinge sequence in repressor function. *J. Bacteriol.*, **176**, 1767–1772.

Cooney,A.J., Tsai,S.Y., O'Malley,B.W. and Tsai,M.-J. (1992) Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTCA response elements, allowing COUP-TF to repress hormonal induction of the vitamin D_3, thyroid hormone and retinoic acid receptors. *Mol. Cell. Biol.*, **12**, 4153–4163.

Crothers,D.M. and Steitz,T.A. (1992) Transcriptional activation by *Escherichia coli* CAP protein. In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 501–534.

de Crombrugghe,B., Busby,S. and Buc,H. (1984) Cyclic AMP receptor protein: role in transcription activation. *Science*, **224**, 831–838.

Ebright,R.H., Ebright,Y.W. and Gunasekera,A. (1989) Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E.coli lac* DNA site. *Nucleic Acids Res.*, **17**, 10295–10305.

Eustance,R.J., Bustos,S.A. and Schleif,R.F. (1994) Reaching out: locating and lengthening the interdomain linker in AraC protein. *J. Mol. Biol.*, **242**, 330–338.

Friedman,A.M., Fischmann,T.O. and Steitz,T.A. (1995) Crystal structure of *lac* repressor core tetramer and its implications for DNA looping. *Science*, **268**, 1721–1727.

Galas,D.J. and Schmitz,A. (1978) DNase footprinting: a simple method for detection of protein–DNA binding specificity. *Nucleic Acids Res.*, **5**, 3157–3170.

Gartenberg,M.R. and Crothers,D.M. (1988) DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature*, **333**, 824–829.

Gerlach,P., Søgaard-Andersen,L., Pedersen,H., Martinussen,J., Valentin-Hansen,P. and Bremer,E. (1991) The cyclic AMP (cAMP)–cAMP receptor protein complex functions both as an activator and as a corepressor at the *tsx-p$_2$* promoter of *Escherichia coli* K-12. *J. Bacteriol.*, **173**, 5419–5430.

Ghosaini,L.R., Brown,A.M. and Sturtevant,J.M. (1988) Scanning calorimetric study of the thermal unfolding of catabolite activator protein from *E.coli* in the absence and presence of cyclic mononucleotides. *Biochemistry*, **27**, 5257–5261.

Herr,W. and Cleary,M.A. (1995) The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev.*, **9**, 1679–1693.

Holst,B., Søgaard-Andersen,L., Pedersen,H. and Valentin-Hansen,P. (1992) The cAMP–CRP/CytR nucleoprotein complex in *Escherichia coli*: two pairs of closely linked binding sites for the cAMP–CRP activator complex are involved in combinatorial regulation of the *cdd* promoter. *EMBO J.*, **11**, 3635–3643.

Kristensen,H.-H., Valentin-Hansen,P. and Søgaard-Andersen,L. (1996) CytR/cAMP–CRP nucleoprotein formation in *E.coli*: the CytR repressor binds its operator as a stable dimer in a ternary complex with cAMP–CRP. *J. Mol. Biol.*, **260**, 113–119.

Kroeger,P.E. and Morimoto,R.I. (1994) Selection of new HSF1 and HSF2 DNA-binding sites reveals differences in trimer cooperativity. *Mol. Cell. Biol.*, **14**, 7592–7603.

Lewis,M., Chang,G., Horton,N.C., Kercher,M.A., Pace,H.C., Schumacher,M.A., Brennan,R.G. and Lu,P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, **271**, 1247–1254.

Liu-Johnson,H.-N., Gartenberg,M.R. and Crothers,D.M. (1986) The DNA-binding domain and bending angle of *E.coli* CAP protein. *Cell*, **47**, 995–1005.

Mavrothalassitis,G., Beal,G. and Papas,T.S. (1990) Defining target sequences of DNA-binding proteins by random selection and PCR: determination of the GCN4 binding sequence repertoire. *DNA Cell Biol.*, **9**, 783–788.

Nagadoi,A. *et al*. (1995) Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. *Structure*, **3**, 1217–1224.

Nørregaard-Madsen,M, Mygind,B., Pedersen,R., Valentin-Hansen,P. and Søgaard-Andersen,L. (1994) The gene encoding the periplasmic cyclophilin homologue, PPIaseA, in *Escherichia coli*, is expressed from four promoters, three of which are activated by the cAMP–CRP complex and negatively regulated by the CytR repressor. *Mol. Microbiol.*, **14**, 989–997.

Pedersen,H., Søgaard-Andersen,L., Holst,B. and Valentin-Hansen,P. (1991) Heterologous cooperativity in *Escherichia coli*: the CytR repressor both contacts DNA and the cAMP receptor protein when binding to the *deoP2* promoter. *J. Biol. Chem.*, **266**, 17804–17808.

Pedersen,H., Søgaard-Andersen,L., Holst,B., Gerlach,P., Bremer,E. and Valentin-Hansen,P. (1992) cAMP–CRP activator complex and the CytR repressor protein bind co-operatively to the *cytRP* promoter in *Escherichia coli* and CytR antagonizes the cAMP–CRP-induced DNA bend. *J. Mol. Biol.*, **227**, 396–406.

Pedersen,H., Dall,J., Dandanell,G. and Valentin-Hansen,P. (1995) Gene regulatory modules in *Escherichia coli*: nucleoprotein complexes formed by cAMP–CRP and CytR at the *nupG* promoter. *Mol. Microbiol.*, **17**, 843–853.

Pollock,R. and Treisman,R. (1990) A sensitive method for the determination of protein–DNA binding specificities. *Nucleic Acids Res.*, **18**, 6197–6204.

Rasmussen,P.B., Søgaard-Andersen,L. and Valentin-Hansen,P. (1993) Identification of the nucleotide sequence recognized by the cAMP–CRP dependent CytR repressor protein in the *deoP2* promoter in *E.coli*. *Nucleic Acids Res.*, **21**, 879–885.

Schumacher,M.A., Choi,K.Y., Zalkin,H. and Brennan,R.G. (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science*, **266**, 763–770.

Smith,D.L. and Johnson,A.D. (1992) A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an α2 dimer. *Cell*, **68**, 133–142.

Søgaard-Andersen,L. and Valentin-Hansen,P. (1993) Protein–protein interactions in gene regulation: the cAMP–CRP complex sets the specificity of a second DNA-binding protein, the CytR repressor. *Cell*, **75**, 557–566.

Søgaard-Andersen,L., Møllegaard,N.E., Douthwaite,S.R. and Valentin-Hansen,P. (1990) Tandem DNA-bound cAMP–CRP complexes are required for transcriptional repression of the *deoP2* promoter by the CytR repressor in *Escherichia coli*. *Mol. Microbiol.*, **4**, 1595–1601.

Søgaard-Andersen,L., Mironov,A.S., Pedersen,H., Sukhodelets,V.V. and Valentin-Hansen,P. (1991a) Single amino acid substitutions in the cAMP receptor protein specifically abolish regulation by the CytR repressor in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **88**, 4921–4925.

Søgaard-Andersen,L., Pedersen,H., Holst,B. and Valentin-Hansen,P. (1991b) A novel function of the cAMP–CRP complex in *Escherichia coli*: cAMP–CRP functions as an adaptor for the CytR repressor in the *deo* operon. *Mol. Microbiol.*, **5**, 969–975.

Valentin-Hansen,P. (1982) Tandem CRP binding sites in the *deo* operon of *Escherichia coli* K-12. *EMBO J.*, **1**, 1049–1054.

Valentin-Hansen,P., Larsen,J.E.L., Højrup,P., Short,S.A., Barbier,C.S. (1986) Nucleotide sequence of the CytR regulatory gene of *E.coli* K-12. *Nucleic Acids Res.*, **14**, 2215–2228.

Valentin-Hansen,P., Søgaard-Andersen,L. and Pedersen,H. (1996) A flexible partnership: the CytR anti-activator and the cAMP–CRP activator protein, comrades in transcriptional control. *Mol. Microbiol.*, **20**, 461–466.

Verrijzer,C.P., Alkema,M.J., van Weperen,W.W., van Leeuwen,H.C., Strating,M.J.J. and van der Vliet,P.C. (1992) The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.*, **11**, 4993–5003.

Vidal-Ingigliardi,D., Richet,E. and Raibaud,O. (1991) Two MalT binding sites in direct repeat: a structural motif involved in the activation of all promoters of the maltose operon in *Escherichia coli* and *Klebsiella pneumoniae*. *J. Mol. Biol.*, **218**, 323–334.

Wade-Jardetzky,N., Bray,R.P., Conover,W.W., Jardetzky,O., Geisler,N. and Weber,K. (1979) Differential mobility of the N-terminal headpiece in the *lac*-repressor protein. *J. Mol. Biol.*, **128**, 259–264.

Weickert,M.J. and Adhya,S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.*, **267**, 15869–15874.

Received on September 20, 1996; revised on December 24, 1996