CRP Binding and Transcription Activation at CRP-S Sites

Andrew D. S. Cameron and Rosemary J. Redfield*

Introduction

The cAMP receptor protein (CRP; also called catabolite activator protein) regulates a global sugar starvation response in three γ-proteobacteria families, the Enterobacteriaceae, Pasteurellaceae, and Vibrionaceae. CRP binds specific sites at gene promoters when activated by its allosteric effector cAMP (a signal of sugar depletion) and then recruits RNA polymerase (RNAP) through direct protein–protein contacts (reviewed in Ref. 2). Because only Escherichia coli CRP has been extensively characterized (reviewed in Refs. 2–4), its structure and function serve as models for understanding CRP in other bacteria. Recent results have also implicated CRP as a regulator of natural competence in several important pathogens, including E. coli, Haemophilus influenzae, Vibrio cholerae, and Salmonella sp.1,5

CRP homodimers bind 22-bp sequences with 2-fold symmetry (consensus half-site 5'-A1A2A3T4G5T6G7A8C10T11) and cause DNA to bend 80–90°.6 CRP's strong preference for the sequence T6G5T4A7G8A8 in both half-sites arises primarily from hydrogen bonds formed between protein side chains and base pairs G5·C, G7·C, and A8·T. In addition, nonclassical hydrogen bonds between the protein and thymine methyl groups at T4·A, T6·A, and A8·T make minor contributions to specificity.7–10 Although CRP contacts the thymine methyl group at T6·A,10 strong selection for T at position 6 arises through an indirect

Received 12 June 2008; received in revised form 11 August 2008; accepted 13 August 2008
Available online 22 August 2008

Received 12 June 2008; received in revised form 11 August 2008; accepted 13 August 2008
Available online 22 August 2008

*Corresponding author. Mailing address: Life Sciences Centre (Zoology), University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia, Canada V6T 1Z3. E-mail address: redfield@zoology.ubc.ca.

Abbreviations used: CRP, cAMP receptor protein; RNAP, RNA polymerase; sBHI, supplemented brain heart infusion.

In Haemophilus influenzae, as in Escherichia coli, the cAMP receptor protein (CRP) activates transcription from hundreds of promoters by binding symmetrical DNA sites with the consensus half-site 5'-A1A2A3T4G5T6G7A8C10T11. We have previously identified 13 H. influenzae CRP sites that differ from canonical (CRP-N) sites in the following features: (1) Both half-sites of these noncanonical (CRP-S) sites have C₆ instead of T₆, although they otherwise have an unusually high level of identity with the binding site consensus. (2) Only promoters with CRP-S sites require both the CRP and Sxy proteins for transcription activation. To study the functional significance of CRP-S site sequences, we purified H. influenzae (Hi)CRP and compared its DNA binding properties to those of the well-characterized E. coli (Ec)CRP. All EcCRP residues that contact DNA are conserved in HiCRP, and both proteins demonstrated a similar high affinity for the CRP-N consensus sequence. However, whereas EcCRP bound specifically to CRP-S sites in vitro, HiCRP did not. By systematically substituting base pairs in native promoters and in the CRP-N consensus sequence, we confirmed that HiCRP is highly specific for the perfect core sequence T₆G₅T₄A₇G₈A₈ and is more selective than EcCRP at other positions in CRP sites. Even though converting C₆ → T₆ greatly enhanced HiCRP binding to a CRP-S site, this had the unexpected effect of nearly abolishing promoter activity. A+T-rich sequences upstream of CRP-S sites were also found to be required for promoter activation, raising the possibility that Sxy binds these A+T sequences to simultaneously enable CRP–DNA binding and assist in RNA polymerase recruitment.

© 2008 Elsevier Ltd. All rights reserved.
readout’ mechanism because the T₆/G₇ base step favors the ∼40° kink needed for DNA deformation by CRP.⁹,¹¹

All characterized *E. coli* CRP sites differ from the consensus at one or more positions, as do the binding sites of other global regulators.¹² Until recently, this has been thought to only reflect selection for stronger or weaker binding, with no significance attached to the presence of specific nonconsensus bases at specific CRP site positions. However, we have found

**Fig. 1.** *H. influenzae* and *E. coli* CRP binding to promoter DNA. (a) Sequence logos of *E. coli* CRP-N sites (*n*=49), *H. influenzae* CRP-N sites (*n*=45), and *H. influenzae* CRP-S sites (*n*=13). The binding site cores (positions 4–8 and 15–19) are highlighted with white columns, and gray arrows at the bottom indicate the inverse palindrome of CRP sites. (b) Alignment of CRP binding sites in promoter DNAs used as bait in bandshift assays. Bases are numbered as in (a), and differences from the binding site consensus (ICAP) are shown. (c–f) Bandshift data. Binding curves were fit to the data except in cases where HiCRP did not bind specifically to promoter DNA. Gray-shaded areas on graphs indicate the amount of nonspecific (ns) DNA binding at high CRP concentrations. (c) HiCRP binding to CRP-N sites. (d) HiCRP binding to CRP-S sites. (e) EcCRP binding to CRP-N sites. (f) EcCRP binding to CRP-S sites. (g) Representative bandshifts from which data in (c) and (e) were derived. (h) Bandshifts illustrating the supershifts characteristic of nonspecific DNA binding.
that the CRP sites of *H. influenzae* competence gene promoters consistently differ from canonical sites in having C₆ and never T₆ in both half-sites. Bioinformatic analysis of other γ-proteobacteria genomes also revealed that the putative CRP sites of Pasteurellaceae, Enterobacteriaceae, and Vibrionaceae competence gene promoters have a consistent over-representation of C₆ suggesting that noncanonical (CRP-S) sites are not an *H. influenzae*-specific phenomenon. However, the functional significance of C₆ in CRP-S sites is unclear because, in *E. coli*, a T₆→C₆ substitution in both halves of a canonical (CRP-N) site reduces CRP–DNA affinity 80-fold.

All *H. influenzae* promoters with CRP-S sites also require the protein Sxy for transcription activation, raising the possibility that Sxy assists CRP binding to CRP-S sites and/or RNAP recruitment. Although the abovementioned three bacterial families all have Sxy homologs, and extensive genetic studies have confirmed Sxy’s role as a regulator of competence genes, direct characterization of its action at CRP-S promoters has been stymied by the toxicity and intractability of this small protein. To better understand CRP binding to CRP-S sites, we have conducted a detailed analysis of *H. influenzae* (Hi) CRP and *E. coli* (Ec) CRP binding and transcriptional activation at native and synthetic CRP sites. Even though the proteins are 78% identical and all of EcCRP’s DNA-binding residues are conserved in HiCRP, the latter is much more selective for binding sites. Nevertheless, two key features are shared: both proteins preferentially bind the same consensus

---

**Fig. 2.** CRP binding to PₗacZ and Pₚila and promoter mutants. (a) Alignment of CRP binding sites as in Fig. 1b. (b–e) Bandshift data plotted as in Fig. 1; PₗacZ and Pₚila binding data from Fig. 1c–f. (f) Representative bandshifts from which data in (c) and (e) were derived.
sequence, 5′-A1A2A3T4G5T6A7A8T9C10T11, and neither protein can activate transcription from *H. influenzae*’s Sxy-regulated promoters in the absence of Sxy, even when bound to promoter DNA.

**Results**

*H. influenzae* and *E. coli* CRP have different affinities for promoter DNA in vitro

Transcriptome and genome analyses in *H. influenzae* have identified 54 CRP-regulated promoters: 41 promoters containing 45 CRP-N sites, which regulate genes for nutrient uptake and central metabolism, and 13 Sxy-dependent promoters containing 13 CRP-S sites. Figure 1a compares a sequence logo of 49 experimentally determined *E. coli* CRP sites (the standard reference set, available at DPInteract) to logos of *H. influenzae*’s CRP-N and CRP-S sites. The *E. coli* and *H. influenzae* CRP-N logos are very similar, showing consistent overrepresentation of the T4G5T6G7A8 motif and the presence of A+T-rich sequences at positions 1 and 2 in both half-sites. The *H. influenzae* CRP-S logo differs from both CRP-N logos in several features: (i) consistent presence of C6 in both half-sites; (ii) stronger core consensus; (iii) substantially stronger consensus at non-core positions, especially at T3 and T9 of both half-sites; and (iv) consistent presence of T rather than A/T at positions 1 and 2 of both half-sites.

To find out if these distinct characteristics of CRP-S sites are responsible for their regulatory differences, we first conducted a detailed analysis of HiCRP binding to CRP-S and CRP-N sites using electrophoretic mobility shift (bandshift) assays; the well-characterized EcCRP was used as a positive control and reference. Figure 1b compares the four CRP sites tested in these assays to the consensus CRP-binding site ICAP. Two of the promoters tested contained CRP-N sites, *H. influenzae* PmglB (mglBAC operon) and *E. coli* PilA (lacZYA operon) and two contained *H. influenzae* CRP-S sites, PpilA (pilABCD operon) and PcomA (comABCDE operon). Binding reactions contained excess nonspecific competitor DNA to approximate in vivo conditions. Because this precluded measurement of equilibrium binding constants (Kobs) or dissociation constants (Kd), for Figs. 1 and 2, we report apparent dissociation constants (Kapp) in terms of active protein. All protein dilution series, DNA preparations, and binding reactions were conducted at least twice and a single curve was fit to all data. Three independent preparations of HiCRP and EcCRP each gave very reproducible binding to PmglB; minor differences between these preparations explain the scatter in the PmglB binding data (Fig. 1c). Tests for nonspecific binding (‘ns’ in Figs. 1, 2, and 3) used three DNA sequences: PpilA with its CRP-S site removed [PpilA(−uni0394)], coding sequence from the *E. coli* hofB gene (as used in Ref. 1), and the pUC19 multiple-cloning site. HiCRP’s highest affinity was for ICAP (Kapp = 1.1 ± 0.1 nM); binding curves are illustrated in Fig. 1c, and a representative bandshift is shown in Fig. 1g. Although PmglB is a very close match to the consensus, HiCRP demonstrated almost 50-fold lower affinity for this promoter (Kapp = 57 ± 7 nM). HiCRP did not bind PilA, PilA, or ComA until protein concentrations were high enough to elicit nonspecific DNA binding (>300 nM) (Fig. 1c and d). Moreover, binding to all

![Fig. 3](image-url). Equilibrium binding constants (Kobs) measured for ICAP and variant sites using bandshift assays. ICAP and variant sites are aligned on the left as in Fig. 1b. Kobs for nonspecific DNA was measured using PpilA(−uni0394). The means and standard errors from replicate experiments are plotted.
three DNAs generated supershifts characteristic of multiple proteins binding to a single DNA molecule (Fig. 1h). The failure of CRP to significantly bind the \( P_{pilA} \) and \( P_{comA} \) sites is quite unexpected, given these promoters’ almost complete dependence on CRP in vivo.\(^{13} \)

EcCRP demonstrated a high affinity for ICAP (\( K_{\text{app}} = 1.0 \pm 0.2 \) nM) equal to that of HiCRP. However, unlike HiCRP, EcCRP also had a high affinity for \( P_{\text{regB}} \) (\( K_{\text{app}} = 1.6 \pm 0.3 \) nM) and \( P_{\text{lacZ}} \) (\( K_{\text{app}} = 6.4 \pm 0.7 \) nM) (Fig. 1e). EcCRP demonstrated lower, but appreciable, affinities for the two CRP-S sites \( P_{pilA} \) (\( K_{\text{app}} = 75 \pm 10 \) nM) and \( P_{comA} \) (\( K_{\text{app}} = 141 \pm 29 \) nM) (Fig. 1f).

Together, these results show that HiCRP and EcCRP differ in at least two ways. First, because \( P_{\text{lacZ}}, P_{\text{pilA}}, \) and \( P_{\text{comA}} \) differ from the consensus at one, two, and three core positions, respectively (Fig. 1b), HiCRP’s 1000-fold preference for ICAP over these sites suggests that the protein is highly selective for the perfect core sequence \( T_G T_G T_G A_S \) in both half-sites. Second, HiCRP greatly prefers the consensus sequence ICAP over \( P_{\text{regB}} \), indicating that HiCRP affinity, unlike EcCRP affinity, is sensitive to bases outside \( T_G T_G T_G A_S \).

**DNA sequence determinants for HiCRP binding to natural CRP sites**

We tested whether the presence of nonconsensus bases at particular core positions was responsible for the poor HiCRP binding to the \( P_{\text{lacZ}} \) and \( P_{\text{pilA}} \) sites by changing these positions to match the CRP-N core consensus (Fig. 2a). In \( P_{\text{lacZ}} \), the nonconsensus base \( T_9 \) was converted to \( A_9 \) to generate \( P_{\text{lacZ-19A}} \); in \( P_{\text{pilA}} \), the CRP-S consensus bases \( C_6 \) and \( G_{17} \) were converted to the CRP-N consensus bases \( T_6 \) and \( A_{17} \) to generate \( P_{\text{pilA-N}} \). Figure 2b and c shows that HiCRP bound with significantly greater affinity to the mutant promoters: \( P_{\text{lacZ-19A}} \) (\( K_{\text{app}} = 1.16 \pm 0.37 \) nM) and \( P_{\text{pilA-N}} \) (\( K_{\text{app}} = 60 \pm 9 \) nM). Further, HiCRP binding to \( P_{\text{lacZ-19A}} \) and \( P_{\text{pilA-N}} \) generated the discrete bands that are characteristic of site-specific binding (Fig. 2f). Thus, in these reaction conditions, HiCRP binding requires that DNA sites perfectly match the consensus \( T_G T_G T_G A_S \) core in both half-sites. Nevertheless, HiCRP’s affinity for \( P_{\text{lacZ-19A}} \) and \( P_{\text{pilA-N}} \) was much lower than that for ICAP despite their having identical perfect-consensus cores, confirming that bases outside of the \( T_G T_G T_G A_S \) core are important for HiCRP binding.

Identical binding experiments revealed that EcCRP has a 5-fold higher affinity for \( P_{\text{lacZ-19A}} \) than \( P_{\text{lacZ}} \) (\( K_{\text{app}} = 1.5 \pm 0.4 \) nM versus \( 8.4 \pm 0.7 \) nM) and a 44-fold higher affinity for \( P_{\text{pilA-N}} \) than \( P_{\text{pilA}} \) (\( K_{\text{app}} = 1.6 \pm 0.3 \) nM versus \( 75 \pm 10 \) nM) (Fig. 2d and e, respectively). Although these mutant promoters still differed from ICAP at six or eight non-core positions, EcCRP exhibited almost identical affinities for them and for ICAP (\( K_{\text{app}} = 1.0 \pm 0.2 \) nM). Thus, when core positions match the consensus, bases outside the core appear to make only a very minor contribution to EcCRP-DNA binding.

**DNA sequence determinants for HiCRP binding to synthetic CRP sites**

Two lines of evidence suggest that positions 3 and 9 (and the reciprocal positions 20 and 14) are important for HiCRP binding. First, these positions have the most significant overrepresentation in the CRP-S logo in Fig. 1a. These are three of the six nonconsensus positions in the low-affinity \( P_{\text{lacZ-19A}} \) site. To systematically address the importance of these positions for HiCRP binding, we constructed variants of ICAP and measured equilibrium binding constants using bandshift assays. These assays used low concentrations of bait DNA and no competitor DNA to allow comparison with established \( K_{\text{obs}} \) data for ICAP.

Figure 3 shows that HiCRP’s affinity for ICAP (\( K_{\text{obs}} = 1.2 \pm 0.2 \times 10^{-10} \) M\(^{-1} \)) was fivefold less than EcCRP’s (\( K_{\text{obs}} = 6.1 \pm 2 \times 10^{-10} \) M\(^{-1} \)); the EcCRP binding constant measured here is consistent with the \( K_{\text{obs}} \) range previously measured using filter-binding assays (\( 4.2 \pm 0.3 \times 10^{-10} \) M\(^{-1} \) to \( 7.1 \times 10^{-10} \) M\(^{-1} \)). Changing ICAP at either \( T_0 \rightarrow C_6 \) or \( A_3 \rightarrow C_9 \) as well as \( T_20 \rightarrow C_20 \) did not significantly change HiCRP affinity. However, simultaneously substituting both \( T_3 \rightarrow C_6 \) and \( A_14 \rightarrow C_{14} \) reduced HiCRP affinity threefold. Thus, HiCRP’s low affinity for \( P_{\text{lacZ-19A}} \) is at least partly due to the nonconsensus bases immediately adjacent to the \( T_G T_G T_G A_S \) cores.

Although CRP binding sites are conventionally treated as 22 bp in length, we have noticed that CRP-S sites are usually flanked by additional A+T-rich sequence, as is ICAP.\(^{21} \) We thus tested whether positions outside the 22-bp consensus sequence contribute to HiCRP binding by converting ICAP’s A+T-rich flanking sequence to G+C-rich sequences; this reduced affinity fivefold.

We also measured binding constants for ICAP variants containing one or both of the CRP-S-like bases \( C_6 \) and \( G_{17} \). A single \( T_8 \rightarrow C_6 \) substitution reduced HiCRP affinity 36-fold, while having both \( C_6 \) and \( G_{17} \) reduced affinity 100-fold. Nevertheless, HiCRP bound the doubly substituted ICAP with a 4-fold greater affinity than nonspecific DNA, indicating that having perfect matches to the consensus binding site at all other positions confers site specificity even in the presence of the unfavorable bases \( C_6 \) and \( G_{17} \).

Unlike HiCRP, binding constants of EcCRP were not significantly affected by substitutions outside of the \( T_G T_G T_G A_S \) core (Fig. 3). Notably, substitutions at positions 6 and 17 did not significantly reduce EcCRP affinity in these assays, although filter-binding assays by Chen et al. showed an 80-fold reduction for the double mutant.\(^{3} \) This unexpected difference is unlikely to be due to our use of a shorter incubation time (20 min rather than 60 min) because the binding we see is unexpectedly high rather than unexpectedly low, as it would be if binding had not reached equilibrium. An alternative explanation may be that the energetics of DNA binding (and perhaps bending) differ between filter and gel assays.
Despite the 96% similarity between EcCRP and HiCRP’s DNA-binding domains, all bandshift data indicated that HiCRP is more selective than EcCRP for DNA sites in vitro. We used SWISS-MODEL to map HiCRP residues on a crystal structure of EcCRP bound to ICAP and found the predicted HiCRP–DNA structure to be completely congruous with the EcCRP–DNA structure (Supplemental Fig. 1). The PROCHECK, PROVE, and WHAT IF algorithms for validating protein structures confirmed that no HiCRP residues are predicted to reshape the DNA-binding domain or to sterically interfere with the protein–DNA interactions that are known to occur in E. coli.

**CRP binding to promoter DNA is insufficient to stimulate transcription in the absence of Sxy**

The inability of HiCRP to bind specifically to P\textsubscript{comA} and P\textsubscript{pilA} in vitro despite its strong regulatory effect in vivo raised the possibility that CRP requires Sxy to facilitate its binding at CRP-S sites. We had two ways to test this hypothesis in vivo: first, by testing whether EcCRP’s intrinsic affinity for CRP-S sites is sufficient for transcription activation and, second, by testing whether HiCRP alone can activate transcription from the P\textsubscript{pilA}\textsuperscript{N} promoter.

EcCRP is known to restore natural transformability to an *H. influenzae* crp null mutant, but the dependence of this complementation on Sxy has not been tested. The *E. coli* crp gene was expressed in *H. influenzae* crp\textsuperscript{-} and sxy\textsuperscript{-} mutants to test whether EcCRP is able to activate the CRP-S sites of *H. influenzae* competence gene promoters in the absence of Sxy. Real-time (quantitative) PCR was used to measure transcription of *comA* 60 min after transfer to the competence-inducing medium MIV, when competence genes are usually maximally expressed. Transformation frequency, which provides the most sensitive assay of competence gene induction, was measured 90 min after cells were transferred. As expected, cells were not transformable in the absence of crp or sxy, but EcCRP restored transformability to a crp\textsuperscript{-} strain (RR815) (Fig. 4a). Also, as previously reported, *comA* expression was induced over 100-fold in MIV, but induction required crp and sxy. However, even though EcCRP binds P\textsubscript{comA}, with high affinity in vitro, it could not restore transformability or *comA* expression in crp\textsuperscript{+}/sxy\textsuperscript{-} cells (RR1130). To confirm that the absence of competence induction was not due to interference between EcCRP and HiCRP in the sxy\textsuperscript{-} strain, we measured transformation in wild-type *H. influenzae* carrying EcCRP (RR1129), and these cells were found to be fully transformable (last column, Fig. 4a). An alternative explanation of the Sxy-dependent competence of cells containing both HiCRP and EcCRP (strain RR1130) would be that all of the EcCRP is sequestered in inactive heterodimers, with the observed expression of competence genes due entirely to active HiCRP homodimers. However, this could only occur if HiCRP were in excess to EcCRP, whereas the reverse is much more likely because in these experiments, crp\textsubscript{E.coli} is plasmid-borne and driven by a constitutive P\textsubscript{lacZ} promoter while crp\textsubscript{*H.influenzae} is chromosomal and driven by its own CRP-repressible promoter.
CRP Binding at CRP-S Sites

The experiments reported in Fig. 2 showed that replacing the CRP-S sites bases C₆ and G₁₇ of PₚilA with their CRP-N-site counterparts (PₚilAN) permitted HiCRP to bind this site in vitro (in the absence of Sxy). To find out whether this change also allowed HiCRP to bind PₚilAN and stimulate transcription in vivo in the absence of Sxy, we used real-time PCR to measure pilAB transcription originating from the promoters of the engineered plasmids. Expression of comA served as a positive control to confirm the presence/absence of the competence-inducing signal in sxy+/− backgrounds, respectively. Because the C₆→T₆/G₁₇→A₁₇ substitutions improved CRP binding in vitro, these changes were not predicted to affect the amount of transcription activation in wild-type cells. Surprisingly, PₚilAN was induced only 3-fold in MIV, significantly less than the 35-fold induction of PₚilA (Fig. 4b). The conversion of C₆→T₆/G₁₇→A₁₇ was hypothesized to relieve Sxy dependence, but even the modest 3-fold induction of PₚilAN was dependent on Sxy. Together, the EcCRP-PₚilAN and HiCRP-PₚilAN results provide complementary evidence that Sxy remains essential for transcription activation in vivo even when CRP can efficiently bind CRP-S-regulated promoters in vitro.

A+T runs upstream of CRP-S sites are required for promoter activation

The above analysis strongly suggested that CRP-S-regulated promoters contain additional elements outside of their CRP-S sites. Previous searches for putative Sxy binding sites in H. influenzae did not identify any conserved motifs other than CRP-S sites, but the search algorithms were insensitive to short (<10 bp) motifs. To detect additional conserved elements in H. influenzae’s CRP-S-regulated promoters, we generated a sequence logo from alignment of the 13 promoter sequences at their CRP-S sites, including 200 bases upstream of predicted transcription start sites. Most CRP-S sites were predicted to be located around −61.5 from transcription start points, and this numbering is used in Fig. 5. The o70−35 and −10 sites apparent in the logo validates our earlier analysis of CRP-S site location and provides strong evidence that competence genes are regulated by o70 in H. influenzae. The sequence logo also revealed striking A+T runs at positions −79, −90, and −102 (runs II, III, and IV), with an additional run in the upstream end of the CRP-S site (run I, at position −71). These A+T runs resemble the RNAP αCTD binding sites called UP elements, contact between αCTD and UP elements enhances transcription from 2- to >100-fold at many bacterial promoters. Although UP elements are usually located between −40 and −60, DNA bending by CRP or IHF has been shown to allow αCTD to bind UP elements at −80 and −90. PₚilA variants in which the A+T runs were replaced or translocated were constructed to test whether these A+T runs are important for promoter activity. As shown in Fig. 5b, replacement of element II (at −79) with a G+C-rich SacII site (CCCGGG)

![Fig. 5. A+T runs upstream of CRP-S sites are important for promoter activation. (a) Sequence logo generated from alignment of H. influenzae’s 13 CRP-S sites. Numbering indicates the average distance of CRP-S sites from predicted transcription start sites identified in Redfield et al. A+T runs are underlined blue and numbered. The sequence logo generated from alignment of 401 E. coli o70 binding sites (copied from Ref. 28) facilitates comparison with similar motifs in the H. influenzae logo above. (b) Schematic of PₚilA mutants and E. coli’s rrnB P1. Fold induction of pilAB after 60 min in MIV is plotted as in Fig. 4b. Baseline pilAB expression in the uninduced state was the same in all strains. (c) Proposed model for the CRP–Sxy–RNAP nucleoprotein complex formed at promoters with CRP-S sites. By binding A+T runs, Sxy could increase DNA curvature, assist CRP binding, and directly assist in RNAP recruitment through contact with αCTD.](image-url)
reduced expression by half. Replacement of element III (at −90) by a SacII site reduced expression by more than eightfold, whereas simultaneous deletion of elements II, III, and IV completely eliminated inducibility. The A+T runs might be important for transcription because CRP-mediated DNA bending allows RNAP’s αCTD subunits to bind them. To test whether the A+T runs are enhancers of RNAP binding, elements II and III were moved to a position immediately adjacent to RNAP’s −35 binding site, thus aligning them with the UP elements of the model UP-enhanced promoter rrrB P134 (Fig. 5b). This translocation did not alter or enhance promoter activity during exponential growth; neither did it allow the promoter to be induced in MIV. Transformation frequencies and comA expression confirmed that competence induction was normal in all strains (data not shown). These results suggest that the A+T runs cannot function as UP elements to recruit RNAP and also show that the CRP-S site is essential for promoter induction.

EcCRP bound normally to the ΔII,III,IV promoter, indicating that the A+T runs are not important for CRP binding (data not shown). Unfortunately, Sxy remains recalcitrant to overexpression and purification; hence, bandshifts could not be used to investigate whether Sxy binds the A+T runs.

Discussion

This study of the molecular mechanisms regulating competence genes in *H. influenzae* is both the first detailed analysis of CRP binding to CRP-S promoters and the first biochemical analysis of CRP from a member of the Pasteurellaceae. Although HiCRP, like EcCRP, demonstrated a high affinity for the consensus CRP site (ICAP), it was much more sensitive than the promiscuous EcCRP to base composition outside the T₆G₅T₇G₈A₉ core. This is surprising given the very high level of similarity of the sequence of its DNA-binding domain to that of EcCRP. Indeed, homology modeling of HiCRP bound to ICAP predicted a protein structure identical with that of EcCRP. The explanation may be that HiCRP and EcCRP are both capable of forming the same protein–DNA contacts, but HiCRP–DNA interactions are more transient because of features in the protein’s less conserved N-terminal domain. For example, HiCRP dimers may be less stable than EcCRP dimers, thus requiring more optimal CRP sites for HiCRP binding in the absence of other cellular proteins. Dimer stability has been previously measured for EcCRP,35 and similar experiments may help explain the DNA-binding characteristics of HiCRP.

Unlike substitutions at other core positions, which eliminate important protein–DNA bonds, the T₆→C₆ substitution inhibits CRP binding by generating a C₆/G₇ base step that increases the free energy required for DNA kinking.9 This unique feature of the T₆→C₆ substitution suggests a distinct mechanistic role for the C₆/G₇ base step in CRP-S function and promoter regulation. *H. influenzae*’s CRP-S sites are distinguished from CRP-N sites not only by having the unfavorable C₆/G₇ base step in both core half-sites but also by having a higher frequency of favorable bases at other core positions and non-core positions. Thus, CRP-S sites appear to have evolved to facilitate targeting of CRP (extensive hydrogen bonding between CRP and DNA) but prevent activation (CRP alone cannot form an activating complex).

DNA topology plays an important role in bacterial promoter regulation (reviewed in Refs. 36–38). The dramatic (~90°) deformation of DNA caused by CRP binding is thought to stimulate transcription both by bringing upstream promoter elements into contact with the polymerase and by facilitating DNA strand separation;239,40 for example, because CRP sites at the *gal* and *malK* promoters can be functionally replaced by intrinsically bent DNA, CRP’s primary role at these sites is to bend DNA.41,42 The inflexibility of CRP-S sites may be specially selected to prevent upstream DNA and/or proteins from interacting with the promoter except in the presence of CRP-induced kinking. However, our results suggest that CRP binding alone cannot stimulate transcription in the absence of Sxy, indicating that CRP-induced kinking alone is insufficient for transcription activation. Furthermore, translocating upstream A+T runs to a position adjacent to the RNAP binding site did not stimulate transcription. Thus, we favor a model in which Sxy binds to the A+T runs and directly assists in RNAP recruitment (Fig. 5c). CRP-induced bending would be essential for this interaction because, in its absence, Sxy bound to A+T runs would be too far upstream to recruit RNAP to the −35 and −10 sites. CRP-S-regulated promoters are unusually strong—global analysis of the *H. influenzae* transcriptome revealed that, in starvation conditions, most promoters with CRP-S sites are induced much more strongly than CRP-N-regulated promoters15—suggesting that, once formed, the putative CRP-Sxy–DNA nucleoprotein complex is very efficient at recruiting RNAP.

We have previously found that EcCRP has very low affinity for its cognate CRP-S promoters.1 Thus, a hallmark of both Enterobacteriaceae and Pasteurellaceae CRP-S sites is that they are low-affinity binding sites for cognate CRP. This low affinity is achieved in a species-specific fashion that corresponds to CRP’s site selectivity. In *H. influenzae*, CRP-S sequences are all strong matches to the CRP-binding site consensus but always include a stiff C₆/G₇ base step that prevents DNA binding by HiCRP in the absence of other proteins. *E. coli* CRP-S sites (which were originally identified in homologs of *H. influenzae* CRP-S-regulated genes) differ from the CRP-binding site consensus at many positions and always have either a C₆/G₇ or a C₆/G₇ base step.43 Thus, CRP-S sites, may serve to create tight regulation without requiring a repressor: basal levels of transcription are very low because the inflexible C₆/G₇ base step limits occupancy by CRP, but expression levels can be very high once activated with the assistance of Sxy.
Materials and Methods

Strains and culture conditions

H. influenzae cells were cultured at 37 °C in supplemented brain heart infusion (sBHI), BHI supplemented with NAD (2 μg/ml) and hemin (10 μg/ml). Novobiocin (2.5 μg/ml), kanamycin (7 μg/ml), or chloramphenicol (2 μg/ml) was added to sBHI when required. H. influenzae strains listed in Table 1 were constructed by transformation of competent cells with chromosomal or plasmid DNA as previously described. To induce competence, exponentially growing (noncompetent) cells were transferred from sBHI to the defined starvation medium MIV. Transformation frequency was measured as the ratio of novobiocin-resistant transformants to total cells after 24 h incubation on sBHI agar with novobiocin. E. coli was cultured in Luria–Bertani broth and made chemically competent with RbCl and transformed as previously described. Chloramphenicol (25 μg/ml) was added when required.

Protein purification and bandshifts

The histidine-tagged CRP proteins were constructed as follows: the crp_Hi, open reading frame was PCR amplified using primers Hi-crp-F and Hi-crp-R (Supplemental Table 1) and cloned in the His-tag vector pQE-30UA (Qiagen) by directly ligating PCR amplicons into the pQE-30UA vector; the crp_E.coli open reading frame was cloned by Peekhaus and Conway in pQE30 (Qiagen) by restriction digestion of crp_E.coli amplicons and the pQE30 vector as described in Ref. 46. His-tagged proteins were expressed and purified as previously described. The fraction of CRP active in sequence-specific DNA binding (usually ~20%) was assessed by titration of ICAP in the absence of competitor DNA.

CRP–DNA binding reactions (5 μl) contained 8 mM Tris (pH 8.0), 30 mM KCl, 3% (v/v) glycerol, 250 μg/ml bovine serum albumin, 100 μM CAMP, and 1 mM dithiothreitol. Reactions were mixed on ice and then incubated at room temperature for 20 min before being loaded onto a 4 °C running polyacrylamide gel (30:1 acrylamide/bisacrylamide; 0.2 × TBE [89 mM Tris, 89 mM borate, and 2 mM ethylenediaminetetraacetic acid (pH 8.3)], 2% glycerol, and 200 μM CAMP; running buffer 0.2 × TBE and 100 μM CAMP). After electrophoresis for 2 h at 10 mA, the gel was dried and exposed (45 min to overnight) to a phosphor screen. Bands were visualized using a STORM 860 scanner (GE Healthcare). Shifted and unshifted radioactivity was quantified using Image Quant (GE Healthcare), and background radioactivity in gel lanes with no CRP was subtracted. Kapp and Kmax values were calculated using Prism 5.0 (GraphPad Software) using nonlinear regression analysis for one-site (specific) binding with Bmax constrained to ≤100% binding.

Bait DNAs were PCR amplified from chromosomal or plasmid templates (primers are listed in Supplemental Table 1) and ranged in size from 90 to 200 bp—these sizes of bait DNA enabled us to accurately measure the amount of bait DNA in bandshift reactions using real-time PCR. Amplicons were purified using polyacrylamide gel electrophoresis; bands were excised and DNA was eluted from macerated gel overnight in TE at 37 °C, followed by ethanol precipitation and resuspension in 10 mM Tris. DNA was end-labeled with T4 polynucleotide kinase using a 10-fold molar excess of γ-[32P]ATP, and unincorporated label was removed using illustra ProbeQuant G-50 Micro Columns (GE Healthcare). Bandshifts contained 35 ng/μl poly(dI–dC) cold competitor DNA and 0.1–2 nM bait DNA (~2100 CPM/fmol) to measure Kapp. For Kmax measurements, protein was in at least 5-fold excess over bait DNA (0.01–0.02 nM).

Table 1. Strains and plasmids used in this work

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant genotype</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H. influenzae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KW20</td>
<td>Wild-type Rd; sequenced strain</td>
<td>43</td>
</tr>
<tr>
<td>MAP7</td>
<td>KW20 Nov</td>
<td>44</td>
</tr>
<tr>
<td>JG57</td>
<td>KW20 crp, Kan’</td>
<td>45</td>
</tr>
<tr>
<td>R864</td>
<td>KW20 xsp: Kan’</td>
<td>15</td>
</tr>
<tr>
<td>R8115</td>
<td>KW20 crp: Kan’ pXN13</td>
<td>27</td>
</tr>
<tr>
<td>R8119</td>
<td>KW20 pXN13</td>
<td>This study</td>
</tr>
<tr>
<td>R8120</td>
<td>KW20 xsp: Kan’ pXN13</td>
<td>This study</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>F80lacZ Δ(lacIYZA-argF) endA1</td>
<td>Qiagen</td>
</tr>
<tr>
<td>M15</td>
<td>pREP4</td>
<td>Qiagen</td>
</tr>
<tr>
<td>R81234</td>
<td>M15 pQEHi-crpr</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Plasmids**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Source or reference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEM-T Easy</td>
<td>Promega</td>
<td></td>
</tr>
<tr>
<td>pQE-30UA</td>
<td>Qiagen</td>
<td></td>
</tr>
<tr>
<td>pREP4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pQEHi-crpr</td>
<td>crp_Hi_influenza in pQE-30UA</td>
<td>This study</td>
</tr>
<tr>
<td>pNP-52</td>
<td>crp_E.coli in pQE30</td>
<td>46</td>
</tr>
<tr>
<td>pSU20</td>
<td>Cam’, F lacZ</td>
<td>47</td>
</tr>
<tr>
<td>pilA</td>
<td>pSU20 containing P_pilA, pilA, and pilB</td>
<td>This study</td>
</tr>
<tr>
<td>pilA2</td>
<td>P_lacZ removed from pilA</td>
<td>This study</td>
</tr>
<tr>
<td>pilA2 ΔIII</td>
<td>A + T run II replaced in pilA2</td>
<td>This study</td>
</tr>
<tr>
<td>pilA2 ΔIII</td>
<td>A + T run III replaced in pilA2</td>
<td>This study</td>
</tr>
<tr>
<td>pilA2 ΔIII, IIIIV</td>
<td>A + T run II, IIIIV removed from pilA2</td>
<td>This study</td>
</tr>
<tr>
<td>pilA2 ΔCRP-S</td>
<td>CRP-S site removed from pilA2</td>
<td>This study</td>
</tr>
</tbody>
</table>

Chromosom region (coordinates 333193–335531) containing ampD, pilA, and the N-terminal half of pilB was PCR amplified and cloned in pGEM-T Easy (Promega) to clone P_pilA. An Accl digest was used to excise a fragment containing P_pilA, pilA, pilB, and 20 bp of the multiple-cloning site. This fragment was cloned in the AccI site in the H. influenzae cloning vector pSU20 to generate plasmid p pilA. H. influenzae does not have the lacI repressor gene, and we found pSU20’s P lacZ to be constitutively expressed; thus, the promoter was removed by an XmnI and XhoI double digest. After DNA was purified on an agarose gel, the sticky end generated by XhoI was filled using the Klenow fragment of DNA Polymerase I and then was ligated to the XmnI-cut blunt end to generate plasmid pilA2.

Site-directed mutations were generated using Strategene’s QuickChange Site-Directed Mutagenesis Kit accord-
ishing to the manufacturer’s instructions (mutagenic primers listed in Supplemental Table 1). The conversion of P<sub>pilA</sub> to P<sub>lacZ</sub>–19A was conducted by site-directed mutation of pSU20. The conversion of P<sub>pilA</sub> to P<sub>lacZ</sub>–N was conducted by site-directed mutation of ppiA2, and all P<sub>pilA</sub> variants listed in Fig. 5 were also constructed in ppiA2. The CRP-S site was removed from ppiA2 using inverse PCR with pfu DNA polymerase and primers ppiA/-35-F and ppiA/ UPII-R, which were designed to bind sequence flanking the CRP-S site. Amplicons contained full ppiA2 sequence except for the CRP-S site and were circularized using blunt-end ligation. The ΔII,III,IV promoter (Fig. 5) was constructed by a SacII digest of p<sub>pilA</sub>2 sequence to bind element II and a second SacII site ∼200 bp upstream; recircularizing the digested plasmid placed vector DNA adjacent to element I.

Real-time (quantitative) PCR

Total RNA was isolated from cultures using RNeasy Mini Kits (Qiagen), and purity and quality were assessed by electrophoresis on 1% agarose (1× TAE). RNA was DNase treated twice with a DNA Free kit (Ambion), and purity and quality were assessed by electrophoresis on 1% agarose (1× TAE). RNA was isolated from cultures using RNeasy Mini Kits (Qiagen), and purity and quality were assessed by electrophoresis on 1% agarose (1× TAE). RNA was DNase treated twice with a DNA Free kit (Ambion), and purity and quality were assessed by electrophoresis on 1% agarose (1× TAE).

References