Escherichia coli cAMP Receptor Protein−DNA Complexes. 1. Energetic Contributions of Half-Sites and Flanking Sequences in DNA Recognition†

Erica A. Pyles,‡ Anita J. Chin, and J. Ching Lee*

Department of Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston, Texas 77555-1055

Received October 2, 1997; Revised Manuscript Received January 23, 1998

ABSTRACT: In *Escherichia coli*, the cyclic AMP receptor protein (CRP) serves as a sensor of the intracellular level of cyclic AMP. At increasing concentrations of cyclic AMP, CRP becomes activated upon binding a cyclic AMP molecule. The activated CRP is capable of regulating the expression of more than 20 genes by binding to specific DNA sites. The specific DNA sequences recognized by CRP consist of two-half-sites of the consensus sequence TGTGA.....XCAXA. At present, the relative contributions of the two half-site and flanking sequences in the energetics of CRP recognition have not been quantitatively defined. A series of 20 DNA sequences was designed to dissect the contributions of individual half-sites and flanking sequences using the natural *gal P1* and *lac P1* sequences as the initial targets. The binding of CRP to these DNA sequences was monitored by fluorescence anisotropy. None of the individual half-sites or flanking sequences contributes more to the binding energetics than a random sequence. In the *lac P1* sequence, the combination of both half-sites leads to a >100-fold increase in affinity compared to that of an individual half-site in CRP−DNA complex formation. The flanking sequence of *lac P1* exhibits a 10- and 0-fold enhancement in affinity for CRP compared to that of a random sequence in the presence and absence of the two half-sites, respectively. The observations of the *gal P1* sequence differ from those of the *lac P1* sequence. The combination of both half-sites exhibits no significant increase in affinity, but the flanking sequence exhibits a 100-fold enhancement in the presence of the two half-sites. However, there is a disproportionate contribution from the flanking sequence proximal to the conserved TGTGA motif. The total energetics of the *gal*−CRP complex formation is essentially due to the presence of the conserved half-site and its adjacent flanking sequence. Thus, the relative contributions of the half-site and flanking sequences to the energetics of DNA recognition are operon specific.

Asymmetry is an important structural feature observed in the regulatory mechanism of gene expression by *Escherichia coli* cAMP receptor protein (CRP),[1] which exists as a stable homodimer and binds to a specific DNA sequence in a 1:1 stoichiometry. Manifestation of asymmetry in CRP−DNA systems can occur in different ways, e.g. variations in the DNA sequences and the structural dynamics of DNA in the DNA−protein complexes. The specific DNA sequences recognized by CRP consist of two half-sites with the consensus sequence TGTGA.....TCACA (*1*, *2*). Few natural sequence variations are observed in the conserved recognition half-site sequence, TGTGA. For example, only one change in the base sequence is observed for the conserved half-site for the *deo PII* and *ara PII* sites (*3*). In contrast, the inverted repeat sequences are not as well conserved as the TGTGA motif. The reason for the divergence from symmetry in the CRP binding sites on DNA is not fully understood. It is known that the perfect palindromic sequence ICAP has an apparent affinity for CRP that is ∼450-fold higher than the apparent affinity of *lac* for the protein (*4*). The dissociation of the ICAP−CRP complex is considerably slower (lifetime of >2 h) than the dissociation of natural CRP binding sites bound to DNA. The slow dissociation of the ICAP−CRP complex may be one reason the natural binding sites diverge significantly in sequence from the palindromic sequence (*5*). A slow dissociation of the CRP−DNA complex would result in either too much or too little gene products and, thus, would hamper the ability of the cell to maintain stasis. Sequences outside of the recognition motifs have a higher probability of base sequence variation than the base sequences found in the recognition motifs. Previous studies have shown that the base sequence located in the flanking ends plays a role in the ability of CRP to induce DNA bending (*6*) and in the energetics of CRP−DNA interactions (*7*). At present, the relative contributions of the two half-sites and flanking sequences in CRP recognition have not been quantitatively defined. Is the TGTGA the primary target for CRP recognition? How much does the flanking sequence contribute to the energetics of the CRP−DNA interaction? Are the

† Supported by NIH Grant GM45579 and Robert A. Welch Foundation Grants H-0013 and H-1238.

* To whom correspondence should be addressed.

‡ Present address: Chemistry Department, Rollins College, Winter Park, FL 32789-4499.

[1] Abbreviations: CRP, cAMP receptor protein; TEDKG buffer, 50 mM Tris, 1 mM K_2EDTA, 1 mM DTT, 100 mM KC1, and 10% (w/v) glycerol at pH 7.80 and 25 °C; TEK(100) buffer, 50 mM Tris, 1 mM K_2EDTA, and 100 mM KC1 at pH 7.8 and 25 °C; TEK(75) buffer, 50 mM Tris, 1 mM K_2EDTA, and 75 mM KC1 at pH 7.80 and 25 °C; TE buffer, 50 mM Tris and 1 mM K_2EDTA at pH 7.80 and 25 °C.

S0006-2960(97)02450-1 CCC: $15.00
Published on Web 03/24/1998

energetic contributions from each of the half-sites and flanking sequences additive?

MATERIALS AND METHODS

Materials

Deprotected and desalted oligodeoxyribonucleotides were purchased from Genosys, or they were synthesized using β-cyanoethyl phosphoramidite chemistry on a Beckman Oligo1000 apparatus. The fluorescent probes 7-(diethylamino)-3-(4′-maleimidylphenyl)-4-methylcoumarin (CPM) and fluroescein 5-maleimide (FM) were purchased from Molecular Probes. The thiol modifier 1-*O*-(dimethoxytrityl)-hexyl disulfide, 1′-[(2-cyanoethyl) (*N*,*N*-diisopropyl)]phophoramidite (C6 S−S), was attached to the 5′-end of the DNA using conventional solid support chemistry. The C6 S−S reagent was purchased from Glen Research and is annotated as *X* in the DNA sequence and name. Other reagents and buffer materials were purchased from Beckman, Sigma, or Boehringer Mannheim.

Methods

Concentrations of solutions employed in this study were determined from the absorption spectrum using the following molar absorption coefficients: CRP monomer at 278 nm (20 400 M^{-1} cm^{-1}; *8*), cAMP at 259 nm (1.465 $\times$ 10^4 M^{-1} cm^{-1}; *9*), and FM at 260, 385, and 494 nm (1.65 $\times$ 10^4, 3.3 $\times$ 10^3, and 7.08 $\times$ 10^4 M^{-1} cm^{-1}, respectively; *10*) in aqueous solution. The molar absorption coefficients for CPM at 260 and 385 nm are 2.3 $\times$ 10^4 and 3.3 $\times$ 10^4 M^{-1} cm^{-1}, respectively, in methanol containing <1% (w/v) β-mercaptoethanol and <1% (w/v) sodium acetate (*11*). Absorption spectra were measured using a Hitachi U-2000 spectrophotometer.

CRP Preparation. CRP was isolated and purified from an overproducing strain (pPLcCRP1) of *E. coli* with the method of Heyduk and Lee (*12*). CRP was stored in TEDKG buffer at −20 °C. Purified CRP was observed to migrate as one band on a sodium dodecyl sulfate−polyacrylamide gel. The integrity of the protein was judged by measuring the stoichiometry and affinity of the protein for the *lac*−CRP binding site in the presence of 200 μM cAMP, and similar results were obtained as previously described using the same solution conditions (*13*). CRP was routinely dialyzed against the desired buffer and gently filtered through a membrane with a pore size of 0.2 μm before being used.

DNA Preparation. The DNA sequences containing the CRP binding sites were synthesized with the recognition sites centrally located in the strand. Each of the sequences flanking the recognition motifs had the same number of nucleotides. A four-letter code is adopted to represent the sequences, as shown in Figure 1. Using this code, the *lac* and *gal* sequences from the 5′- to 3′-end are represented by *LCLL* and *GCGG*, respectively, where *C* is the conserved TGTGA motif. The modified *GCLG* and *XGCCG* sequences are essentially the *gal*−CRP binding site with the inverted repeat ATAAA replaced with the inverted repeat of the *lac*, TCACT, and the perfect palindromic sequence TCACA, respectively. *LCGL* is the DNA sequence that contains the recognition half-sites of *gal P1* substituted for the *lac P1* half-sites, and the DNA sequence *RCGR* has the *gal P1*

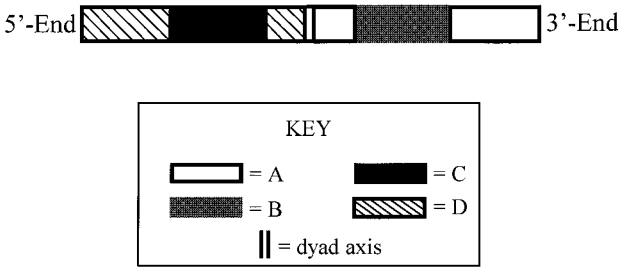

FIGURE 1: Cartoon of CRP binding sites. Specific regions within the CRP binding site are highlighted and annotated in the figure: (A) flanking sequence distal to the conserved TGTGA motif, (B) inverted repeat motif, (C) conserved TGTGA motif, and (D) flanking sequence proximal to the conserved TGTGA motif.

recognition motifs embedded in the *random* sequence. The sequences named *RCRR*, *RRLR*, and *RCLR* contain either one or both of the recognition half-sites of the *lac*−CRP binding site, and these motifs are embedded in the *random* sequence. Substitution of the recognition motifs of *lac P1* for random sequences yields the *LRRL* sequence. Various substitutions of the *gal P1* sequence are listed in Table 1 in the four-letter code with their respective sequence. Each sequence is further assigned a number.

Covalent attachment of CPM to the thiol-modified single-stranded (ss) DNA (*GCCGX* and *LCLLX*) was performed prior to purification of the ssDNA to reduce the extent of oxidation of the disulfide. For the *XGCCG* and *XLCLL* sequences, CPM was covalently attached via the thiol modifier C6 S−S after reducing the disulfide bond with dithiothreitol as previously described (*13*, *14*). Subsequently, the CPM-labeled sequence was purified by denaturing polyacrylamide gel electrophoresis (15% acrylamide and 7.5 M urea) in the dark to separate nonfluorescent ssDNA from fluorescent ssDNA. Sequences that do not contain the disulfide linker were purified by denaturing polyacrylamide gel electrophoresis prior to covalent labeling with the fluorophore (CPM or FM) as previously described (*13*, *14*). Each oligomer was then hybridized to its complementary strand. The integrity of the oligomers was monitored by nondenaturing polyacrylamide gel electrophoresis (15% acrylamide), and each oligomer migrated as one band. No fluorescent ssDNA was observed, and if ssDNA was present, it constituted $\leq$5% of the intensity of the bands observed. The probe:DNA ratio (singly labeled DNA) was calculated from the absorption spectrum of the double-stranded, fluorescently labeled DNA, and it ranged from 0.4 to 1.0 mol of probe per mole of DNA. The doubly labeled DNA contained equimoles of donor and acceptor fluorophores covalently attached to their 5′-ends (0.65 mol of CPM:0.65 mol of FM:1 mol of DNA). The samples were dialyzed against TE buffer and stored at −20 °C.

Fluorescence Measurements. Spectral properties of fluorescently labeled DNA were measured using a SLM 8000C spectrofluorometer equipped with Glan-Thompson calcite prism polarizers and a circulating water bath at 25.0 $\pm$ 0.1 °C.

The anisotropy of CPM−DNA was measured with the spectrofluorometer operated in the T-format with analogue detection. The excitation wavelength was 390 nm, and an Oriel band-pass filter (model 59816) was placed in the excitation path to reduce second-order scatter. Oriel band-pass filters (model 59850) (>25% transmittance between 470

Table 1: Association Constants for CRP−DNA Complexes[a]

no.	CRP binding site	sequence[b]	K_{app} x 10^{-6} (M^{-1})[c]
S1	*RRRR* (*random*)	CTCAGTTCTGATACCAAGCAGCCCAG	0.045 (0.017, 0.073)[d,e]
S2	RCRR	CTCAG**TGTGA**ATACCAAGCAGCCCAG	0.065 (0.055, 0.068)
S3	RRLR	CTCAGTTCTGATACCA**TCACT**CCCAG	0.040 (−0.010, 0.091)
S4	*LCLL* (*lac PI*) →	ATTAA**TGTGA**GTTAGC**TCACT**CATTA	41 (35, 49)[d]
S5	*XLCLL* →	XATTAA**TGTGA**GTTAGC**TCACT**CATTA[f]	54 (39, 74)
S6	LRRL	ATTAATTCTGGTTAGCAGCAGCATTA	0.034 (0.012, 0.055)
S7	RCLR	CTCAG**TGTGA**ATACCA**TCACT**CCCAG	5.3 (4.4, 6.2)
S8	LCGL	ATTAA**TGTGA**GTTAGC**ATAAA**CATTA	0.95 (0.77, 1.1)
S9	*GCGG* (*gal PI*) ←	AAAAG**TGTGA**CATGGA**ATAAA**TTAGT	9.5 (8.5, 11)[d]
S10	GCLG	AAAAG**TGTGA**CATGGA**TCACT**TTAGT	160 (130, 200)[d]
S11	XGCCG	XAAAAG**TGTGA**CATGGA**TCACA**TTAGT	2500 (960, 17100)
S12	RCGG	CTCAG**TGTGA**CATGGA**ATAAA**TTAGT	3.2 (2.9, 3.6)
S13	GRGG	AAAAGTTCTGCATGGA**ATAAA**TTAGT	0.13 (0.11, 0.16)
S14	GCRG	AAAAG**TGTGA**CATGGAAGCAGTTAGT	17 (14, 19)
S15	GCGR	AAAAG**TGTGA**CATGGA**ATAAA**CCCAG	11 (10.5, 12)
S16	RCGR	CTCAG**TGTGA**ATACCA**ATAAA**CCCAG	0.10 (0.081, 0.12)
S17	GRRG	AAAAGTTCTGCATGGAAGCAGTTAGT	0.084 (0.072, 0.095)
S18	RRGG	CTCAGTTCTGATAGGA**ATAAA**TTAGT	0.081 (0.066, 0.096)
S19	GCRR	AAAAG**TGTGA**CATCCAAGCAGCCCAG	5.8 (4.7, 6.9)
S20	gal-16/-42	TAACAAAGATGCGAAAAGTGTGACAT	0.082 (0.07, 0.10)

[a] Measured in TEK(100) in the presence of 200 μM cAMP at pH 7.80 and 25.0 °C. [b] Sequences are listed from the 5′-end to the 3′-end. The recognition half-sites are bold, and *random* sequences are underlined. Arrows represent the direction of transcription. [c] Apparent equilibrium constants of CRP binding to DNA were calculated from eq 3 in Materials and Methods. [d] From ref *7*. [e] Errors in parentheses are expressed in terms of 75% confidence intervals. [f] The letter *X* represents the thiol linker covalently attached to the 5′-end via conventional solid support chemistry.

and 570 nm) were used in the emission paths. The dimensions of the quartz cells used were 10 × 10 mm. CPM−DNA (12 nM) solutions were titrated with CRP in the presence of cAMP (200 μM) in TEK(100) buffer at pH 7.8 and 25.0 °C. The following reaction best describes the global binding of CRP to DNA:

$$PA_n + D \overset{K_{app}}{\rightleftharpoons} PA_nD \quad (1)$$

$$K_{app} = \frac{[PA_nD]}{[PA_n][D]} \quad (2)$$

where D, PA_n ($n = 0$, 1, or 2), and PA_nD are DNA, CRP−cAMP, and CRP−cAMP−DNA, respectively. $P_T = PA_n + PA_nD$. K_{app} is the apparent equilibrium constant. The affinity of CRP for DNA in the absence of cAMP is very weak (*7*). Thus, the contribution of the CRP−DNA complex to the total CRP concentration was neglected. The following equation results after substituting the total protein concentration, P_T, into eq 2 and accounting for all species with anisotropic motion:

$$r_{obs} = r_o + \Delta r\left[\left[K_{app}D_T + K_{app}P_T + 1 - \sqrt{(K_{app}D_T + K_{app}P_T + 1)^2 - 4K_{app}{}^2D_TP_T}\right]/2K_{app}D_T\right] \quad (3)$$

r_{obs}, r_o, Δr, P_T, and D_T are the observed anisotropy, the anisotropy of free DNA, the total change in anisotropy, and the total protein and DNA concentrations, respectively. Dilution of solutions during the titration affected the fitted parameters when the change in volume was >6%. Therefore, the effect of dilution was taken into consideration during data analysis when the volume change was >6%. At least two titrations were performed for each CRP−DNA interaction examined, and the curves were fit simultaneously to eq 3 using the software SCIENTIST (MicroMath Scientific Software, Salt Lake City, UT). The fitted parameters were obtained by nonlinear curve fitting. Errors are expressed in terms of 75% confidence intervals (support plane method).

RESULTS

To define the contribution of half-sites and flanking sequences involved in the energetics of CRP recognition, the apparent equilibrium constant for CRP binding to DNA was quantitatively examined by monitoring the change in anisotropy of CPM-labeled DNA induced by binding to CRP (*13*). CRP binding to DNA was measured as a function of DNA sequence in the presence of 200 μM cAMP in TEK-(100) buffer at pH 7.8 and 25.0 °C (see Table 1 for sequences). The binding isotherms of CRP binding to *RCLR* (S7), *RCRR* (S2), and *RRLR* (S3) are shown in Figure 2. The binding isotherms were fit to eq 3 in Materials and Methods, and the apparent equilibrium constants are listed in Table 1. The random distribution of the residuals is an indication of the goodness of fit of the data. The previously determined apparent equilibrium constants for CRP binding to *LCLL (lac)*, *GCLG (gallac)*, and *GCGG (gal)* under the same conditions are also listed in Table 1 for comparison (*7*).

The binding constant of CRP−RRRR (*random* sequence, S1) is very weak and serves to reflect on the affinity of CRP for a nonspecific sequence (*7*). Sequences *RCRR* (S2) and *RRLR* (S3) are designed to evaluate the contribution from each half-site in the *lac* sequence with respect to specific DNA recognition by CRP. The apparent equilibrium constants for CRP binding to the *RRRR* (S1), *RCRR* (S2), and *RRLR* (S3) are the same. These results are surprising and indicate that a sole half-site in the recognition motif (i.e. the conserved TGTGA or the inverted repeat) within a CRP binding site is not sufficient to obtain specific DNA binding to the protein.

Embedding two specific recognition half-sites (those of *gal* or *lac*) within the random sequence was performed to test the ability of CRP to recognize the DNA. Placement

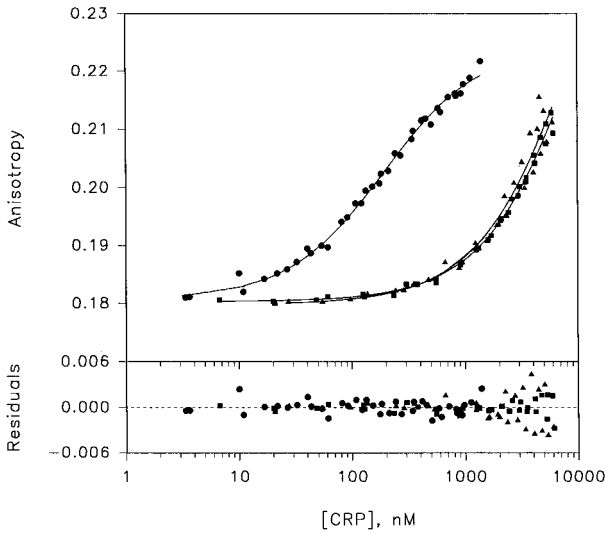

FIGURE 2: Typical binding isotherms of CRP binding to DNA sequences in the presence of 200 μM cAMP in TEK(100) buffer at pH 7.80 and 25.0 °C. The anisotropy of CPM-labeled DNA was normalized to 0.1800. The symbols correspond to the following sequences: *RCLR* (●), *RCRR* (■), and *RRLR* (▲). The length of the CRP binding sites was 26 base pairs. Data were fit to eq 3 (in Materials and Methods), and at least two binding curves were fit simultaneously for each DNA sequence. The goodness to fit is shown by the random distribution of residuals about zero.

of the two recognition half-sites of the *lac P1* within the *RRRR* sequence yields *RCLR* (S7). Binding of *RCLR* (S7) to CRP is significantly increased relative to that of the *random*−CRP complex. These results demonstrate that both half-site motifs are necessary for a specific interaction between *lac P1* and CRP. Surprisingly, the conclusion derived from the interaction between *lac P1* and CRP does not apply to that between *gal P1* and CRP. Placement of both recognition half-sites of the *gal P1* within the random sequence (*RCGR*, S16) does not significantly increase the strength of the interaction. The value of K_{app} is barely greater than the higher limit of that determined for *RRRR* (S1) binding to CRP. The observations of CRP binding to the *lac P1* and *gal P1* and their respective derivatives, *RCLR* (S7) and *RCGR* (S16), imply that the major contribution to DNA recognition does not necessarily reside in the two half-sites identified by statistical analysis. The contributions of the flanking sequences are tested with the sequences *LRRL* (S6) and *GRRG* (S17). Within experimental uncertainties, the values of K_{app} for CRP−*LRRL* and CRP−*GRRG* complexes are no different from that of the *RRRR*−CRP complex. Thus, the flanking sequences alone in either the *lac P1* or *gal P1* sequence do not provide enough information for the protein to distinguish these sequences as specific binding sites. These results suggest that specific DNA sequence recognition by CRP is not the consequence of additive effects of separate components of the DNA sequence.

The value of K_{app} for the CRP−*RCLR* (S7) complex is greater than that for the CRP−*RCGR* (S16) complex, implying that the inverted repeat half-site sequence L contributes more than G in DNA recognition. A series of sequences were designed to test the validity of this conclusion. *GCLG* (S10) exhibits higher affinity for CRP than *GCGG* (S9), while *LCLL* (S4) has a higher affinity for CRP than *LCGL* (S8). Thus, regardless of the sequence context, the inverted repeat half-site sequence L contributes more than G in DNA recognition. It is interesting to note that a placement of C in the inverted repeat half-site significantly enhances the affinity of CRP for DNA as shown by the interaction of CRP with *GCCG* (S11). Hence, it may be concluded that, within the *gal P1* sequence of *GCYG*, the contribution of Y to DNA recognition increases in the order G < L < C.

Substitution of the *CC* recognition motifs (perfect palindrome) for those of *gal P1* into the *gal P1* background (*GCCG*, S11) yielded a ssDNA sequence that has a greater propensity to form hairpin structures in solution. Thus, the fluorescence labeling method was modified. As a control, *lac PI* was labeled in a similar manner. Attachment of the fluorophore through the thiol modifier did not alter the apparent affinity of CRP for DNA since similar apparent equilibrium constants for *LCLL* (S4) and *LCLL* (S5) binding to CRP were observed (Table 1). However, the increased spacer length decreased the initial anisotropy of CPM−DNA to 0.1605 ± 0.001 compared to the values for the other CPM−DNA sequences employed in this study (0.1872 ± 0.0068; *7*). Nevertheless, the overall change in anisotropy of CRP−*XLCLL* and CRP−*XGCCG* complexes was unaffected by the larger spacer length between CPM and the end of DNA, and the total change in anisotropy was similar to those observed for CRP−*LCLL* and CRP−*GCLG* (*7*). CPM−*XGCCG* binding to CRP is very tight, and the interaction is stoichiometric in the presence of 100 mM KCl (data not shown). Therefore, the value listed in Table 1 is a lower estimate of the apparent equilibrium constant of CRP−*XGCCG* interaction with 100 mM KCl.

The contribution of each DNA segment, as specified in Figure 1, to the affinity of CRP for the wild type *gal P1* sequence is tested by systematically randomizing one segment at a time. *RCGG* (S12) and *GCGR* (S15) are designed to evaluate the contribution of the flanking sequence proximal and distal, respectively, to the conserved TGTGA (C) motif. It is interesting to note that the value of K_{app} for the CRP−*RCGG* (S12) complex is significantly lower than that of the *gal P1*−CRP complex (*GCGG*, S9), while the corresponding value for the CRP−*GCGR* (S15) complex is essentially identical to that for the CRP−*GCGG* complex. These results imply that there is a disproportionate contribution from the flanking sequences with respect to DNA recognition by CRP. In *gal P1*, the flanking sequence proximal to the conserved TGTGA motif contributes significantly to the protein−DNA interaction, while the distal one does not.

The contribution of each half-site in the context of the *gal P1* sequence is tested. Randomizing the conserved TGTGA motif, as represented by *GRGG* (S13), significantly lowers the affinity of CRP for this derivatized *gal* sequence. However, randomizing the inverted repeat half-site, as represented by *GCRG* (S14), shows no effect on the value of K_{app}. That is, the apparent affinities of CRP for *GCGG* (*gal P1*, S9) and *GCRG* (S14) are the same. Thus, like the observation on the flanking sequences in *gal P1*, there is a disproportionate contribution by the two recognition half-sites. As a further test of this phenomenon of disproportionate contribution, the sequences of *RRGG* (S18) and *GCRR* (S19) were tested. Randomizing the conserved TGTGA motif and its adjacent flanking sequence (*RRGG*, S18) reduced the value of K_{app} for the protein−DNA interaction to that of the *random* sequence (S1), whereas there is no

observable effect on K_{app} caused by randomizing the other two segments of the DNA sequence (*GCRR*, S19).

DISCUSSION

CRP regulates more than 20 genes in *E. coli*. Hence, a major issue in elucidating the molecular mechanism of CRP action is the determination of factors involved in specific DNA sequence recognition. What are the contributions of the two half-sites in the energetics of CRP−DNA interaction? Does the highly conserved TGTGA half-site contribute more? What is the role of the sequence flanking the two half-sites? Another important feature in CRP action is the induced bending of the DNA binding site. Does CRP bend all of these specific sites symmetrically? These are the issues being addressed.

The apparent association constants for CRP binding to the *RCRR* (S2) and *RRLR* (S3) sequences are indistinguishable from that of the *random* sequence (*RRRR*, S1) (Table 1). These results imply that each individual half-site does not contribute significantly to the energetics of recognition between CRP and a specific DNA sequence. These results are surprising since the *RCRR* (S2) sequence contains the highly conserved TGTGA motif. Furthermore, the *LRRL* (S6) and *GRRG* (S17) sequences do not exhibit affinities any higher than that of CRP binding to *RRRR*. These results demonstrate that the sole sequences identified as recognition half-sites by statistical analysis and the flanking sequences do not provide CRP a means for initial recognition of a specific binding site on DNA. However, the *RCLR* (S7) sequence, which consists of both specific half-sites in *lac P1* identified by statistical analysis, exhibits an affinity for CRP >100-fold greater than that of the *RRRR* sequence for CRP. This result implies that both half-sites must be present in order to obtain high specificity between the DNA and the protein. Thus, the binding of CRP to specific DNA sites consists of a nonadditive energetic contribution from each of the half-sites. Nevertheless, it is important to remember that these half-site motifs are not always required for specific CRP−DNA interaction. This was clearly demonstrated in previous thermodynamic measurements that show *crp PII* binding to the protein has the same affinity as *crp P1* binding to CRP, even though the base sequence in the former binding site does not contain the recognition half-sites whereas the latter binding site does (*7*).

Knowledge of the apparent equilibrium constants allows for the calculation of the change in free energy associated with the reaction ($\Delta G_t = -RT \ln K_{app}$), as summarized in Table 2. Measurement of the contribution of the two recognition half-sites placed together or by themselves in the *random* sequence (*RRRR*, S1) enabled us to dissect the energetic contributions of nonspecific and specific protein−base interactions to the total change in free energy. Assuming

$$\Delta G_t = \Delta G_s + \Delta G_{ns} \quad (4)$$

and

$$\Delta G_s = \Delta G_i + \Delta G_f \quad (5)$$

therefore,

Table 2: Changes in Free Energy of CRP−DNA Complex Formation and Contribution of Recognition Motifs to the Total Free Energy of CRP Binding to DNA[a]

DNA sequence	$-\Delta G^b$ (kcal/mol)	$-\Delta G_s$ (kcal/mol)	$-\Delta G_i$ (kcal/mol)	$-\Delta G_f$ (kcal/mol)
S1, *RRRR*	6.4 (0.4)[c,d]			
S2, *RCRR*	6.6 (0.1)			
S3, *RRLR*	6.3 (0.7)			
S4, *LCLL*	10.4 (0.1)[d]	4.0 (0.4)	2.8 (0.4)	1.2 (0.6)
S6, *LRRL*	6.2 (0.5)			0.2 (0.6)
S7, *RCLR*	9.2 (0.1)	2.8 (0.4)	2.8 (0.4)	
S8, *LCGL*	8.2 (0.1)	1.8 (0.5)	0.5 (0.5)	1.2 (0.6)
S9, *GCGG*	9.5 (0.1)[d]	3.1 (0.4)	0.4 (0.5)	2.7 (0.6)
S10, *GCLG*	11.2 (0.1)[d]	4.9 (0.5)	2.8 (0.4)	2.1 (0.6)
S11, *XGCCG*	12.8 (0.9)	6.5 (1.0)	3.8 (1.2)	2.7 (0.6)
S12, *RCGG*	8.8 (0.1)	2.4 (0.4)		
S13, *GRGG*	7.0 (0.1)	0.6 (0.4)		
S14, *GCRG*	9.8 (0.2)	3.4 (0.5)		
S15, *GCGR*	9.6 (0.1)	3.2 (0.4)		
S16, *RCGR*	6.8 (0.2)	0.4 (0.4)		
S17, *GRRG*	6.7 (0.2)	0.3 (0.4)		
S18, *RRGG*	6.7 (0.1)	0.3 (0.4)		
S19, *GCRR*	9.2 (0.1)	2.8 (0.4)		
S20, *gal-16/-42*	6.7 (0.3)	0.3 (0.5)		

[a] K_{app} was measured in TEK(100) buffer and 200 μM cAMP at pH 7.8 and 25.0 °C. [b] $\Delta G_t = -RT \ln K_{app} = \Delta G_s + \Delta G_{ns} = \Delta G_f + \Delta G_i + \Delta G_{ns}$, where $\Delta G_{ns} = \Delta G_t$ (*random*) = −6.4 kcal/mol. [c] Standard deviation in parentheses. [d] From ref *7*.

$$\Delta G_t = \Delta G_i + \Delta G_f + \Delta G_{ns} \quad (6)$$

ΔG_t is the total free energy change observed in CRP binding to a specific DNA sequence; ΔG_{ns} and ΔG_s are the contributions of nonspecific and specific interaction free energies for CRP−DNA complex formation, respectively. ΔG_{ns} is equivalent to the change in free energy of the *random*−CRP interaction. Thus, ΔG_s is equal to the difference between the change in free energy of a specific CRP−DNA interaction and the change in free energy of *random*−CRP complex formation. The change in free energy due to specific protein−DNA interactions (ΔG_s, direct readout) is a sum of the change in free energies due to the interaction of the two recognition half-sites with CRP (ΔG_i) and the change in free energy due to the interaction of CRP with the flanking sequence (ΔG_f).

The averaged value of ΔG_t for CRP binding to *LRRL* (S6), *RCRR* (S2), and *RRLR* (S3) is −6.4 ± 0.2 kcal/mol, which is identical to that for *random*−CRP interaction at −6.4 ± 0.4 kcal/mol. Hence, ΔG_s values for those sequences are zero, in accordance to eq 4, since ΔG_{ns} is assumed to be equivalent to that for *random*. The simultaneous presence of both half-sites, as represented by *RCLR* (S7), has a synergistic effect on the binding energetics. $\Delta G_s = -2.8$ kcal/mol, and ΔG_s contains no energetic contribution from the flanking sequence because it is part of the nonspecific background sequence whose contribution is assumed to be accounted for by ΔG_{ns}. The source of this substantial increase in binding energetics must include, at least in part, entropy. The interaction between each half-site with CRP can be considered a univalent−univalent reaction. When both half-sites are present in a covalently linked sequence, the protein−DNA interaction could be considered a bivalent−univalent reaction. In any reaction, there is a contribution of a cratic term to ΔG (*15*) and the magnitude is related to

$$RT \ln \frac{X_c}{X_a X_b} \quad (7)$$

where X_a, X_b, and X_c are mole fractions of reactants a and b and product c, respectively. Using the standard state of 1 *m* and considering the reaction in dilute aqueous solution at 25 °C, the cratic contribution amounts to 2.4 kcal/mol, i.e. an unfavorable loss of entropy in the formation of a complex from two separate reactants. This is the unfavorable contribution in a univalent−univalent reaction, such as in the interaction of CRP with a single half-site. A similar contribution is in the initial formation of a complex between a univalent and a bivalent reactant, such as in the association of CRP with a DNA sequence containing two half-sites (CRP-2 half-site). However, in formation of a noncovalent complex between CRP and the second half-site, there is no contribution of the unfavorable entropy term because there is no change in mole fractions of reactants. As a consequence, the energetics of formation of a bivalent−univalent complex should be stronger by 2.4 kcal/mol than the arithmetic summation of the separate univalent−univalent reactions. It is interesting to note that the expected contribution of a cratic term can essentially account for the observed ΔG_s (−2.8 kcal/mol) in *RCLR*−CRP complex formation. Thus, the nonadditivity in the energetics of half-site−CRP interaction in *lac P1* is most likely the simple consequence of having both half-sites covalently linked to a single sequence. This conclusion is consistent with the report that introduction of a single strand nick at the center of a symmetrized *gal* site reduces the binding energy of CRP by about 0.6 kcal/mol (*16*).

The observation for the *gal P1* (S9) is different. The energetic contribution of the two half-sites, CG, amounts to only −0.4 kcal/mol (Table 2). This result implies that the sequence ATAAA contributes less to the energetics of DNA recognition by CRP than TCACT, the equivalent sequence in *lac P1* (S4). This conclusion is supported by comparing the values of ΔG_t for the interactions of CRP with *GCGG* (S9), *GCLG* (S10), and *GCCG* (S11). The DNA sequences (S9, S10, and S11) differ from each other only in the inverted repeat half-site sequence (see Figure 1 and Table 1). Binding of CRP to *GCCG* is about 1.6 kcal/mol more favorable than binding to *GCLG* which in turn is about 1.8 kcal/mol more favorable than binding to *GCGG*. A change in the background sequence to *lac P1* does not alter the conclusion that the sequence TCACT exhibits an affinity higher than that of ATAAA as is evident in the results for S7 and S8. The same conclusion can be derived from comparing the values of ΔG_t for *GCGG* (S9) and *GCRG* (S14). There is no difference in the value of ΔG_t when the sequence ATAAA is replaced by a *random* sequence. Does the flanking sequence proximal to ATAAA have an effect on the energetics of recognition of this half-site sequence? Within experimental uncertainties, the ΔG_t value for *GCRR* (S19) is essentially the same as that for *GCGG* (S9). Thus, the flanking sequence proximal to the ATAAA sequence does not provide any influence on this half-site to the energetics of binding. So, is this sequence essential for CRP binding? The S20 sequence is designed so that the total length of 26 base pairs is maintained; however, the sequence is shifted so as not to include the inverted repeat half-site and the flanking sequence adjacent to it. The binding result shows that this half-site is essential because elimination of this sequence, as represented by the S20 sequence shown in Table 1, results in a low binding affinity that is similar to that observed for the interaction of CRP with the random sequence. Why is the cratic contribution observed in the *lac* half-sites not observed in *gal*? One possible reason is that in *lac P1*, both half-sites are bound to the DNA recognition F helix in CRP and the DNA is bent symmetrically (*14*). Whereas in the *gal* site, the bending may not be symmetrical and the half-sites are not recognized simultaneously. The structure of DNA bound to CRP is monitored and discussed in the following paper (*17*).

Having determined the contribution of both half-sites in the *lac* and *gal* sites, we were able to estimate the contribution of the flanking sequence in the *lac* and *gal* sites. ΔG_t of *lac*−CRP is −10.4 kcal/mol. According to eq 4, ΔG_s is −4.0 kcal/mol. Since ΔG_i is −2.8 kcal/mol, in accordance with eq 5, ΔG_f is −1.2 kcal/mol. Using the same rationale, the contribution of the flanking sequence in the *gal* site can be estimated. ΔG_t of *RCGR*−CRP is −6.8 kcal/mol, and ΔG_i for *gal*−CRP is −0.4 kcal/mol, which corresponds to the energetic contribution of both half-sites in the *gal* site to ΔG_t. Knowing ΔG_t for *gal*−CRP is −9.5 kcal/mol, we find that ΔG_s is −3.1 kcal/mol, in accordance with eq 4. By applying these values and $\Delta G_i = -0.4$ kcal/mol to eq 5, we find that $\Delta G_f = -2.7$ kcal/mol, which corresponds to the energetic contribution of the flanking sequence in the *gal* site to the total free energy of the CRP−*gal* interaction. Results of this analysis indicate that the relative contribution in the energetics of specific DNA site recognition is operon specific. In the *lac* operon, the combination of both half-sites contributes twice as much energetically toward DNA−CRP complex formation as the flanking sequence (−2.8 vs −1.2 kcal/mol). However, the opposite is observed in the *gal* operon. The flanking sequence contributes more than the combination of both half-sites (−2.7 vs −0.4 kcal/mol). One must recognize that the above analysis and conclusion are derived from data in the context of a complete CRP binding sequence in *lac P1* or *gal P1*. When the flanking sequences are taken out of the context of a complete CRP binding sequence, as represented by *LRRL* (S6) and *GRRG* (S17), there is essentially no energetic contributions from these flanking sequences to the CRP−DNA interaction. These results imply that there is communication between the recognition half-sites and flanking sequences. A prime example is the result for the *GCRR* (S19) sequence. The value of ΔG_s for the interaction of CRP with that sequence is −2.8 kcal/mol. Again, one may invoke the cratic entropy contribution to rationalize for the observed value of −2.8 kcal/mol. Thus, the sequence that contributes significantly to the energetics of DNA recognition is operon dependent.

Having dissected the energetic contributions of half-sites and flanking sequences, we were able to test the predictive power of these values and the concept of energetic additivity. An adherence to additivity implies that there is no communication between the half-sites and flanking sequences. Knowing the values of ΔG_f (*lac*) and ΔG_i (*gal*), we found that the predicted value for ΔG_s for *LCGL* (S8) is $-1.2 + (-0.4) = -1.6$ kcal/mol. This value is in excellent agreement with the experimentally determined value of −1.8 kcal/mol (Table 2). A similar calculation predicts a value of −5.5 kcal/mol for ΔG_s of the CRP−*GCLG* complex (S10).

Within experimental error, this value is similar to the experimental value of −4.9 kcal/mol. Thus, within the limited number of cases tested, the two half-sites together and the flanking sequences behave independently in contributing to the energetics of DNA recognition by CRP in the context of a complete CRP binding sequence.

The observation of differential energetic contributions by the flanking sequences and the half-sites in DNA recognition may be linked to the extent of CRP-induced bend in *gal* and *lac* since bending DNA requires an energetic cost (*18*, *19*). In the following paper (*17*), fluorescence spectroscopy was employed to monitor the symmetry of DNA bending in order to establish a correlation between the energetics of binding and the structure of bent DNA in a DNA−CRP complex.

REFERENCES

1. Ebright, R. H., Cossart, P., Gicquel-Sanzey, B., and Beckwith J. (1984) *Nature 311*, 232−235.
2. Berg, O. G., and von Hippel, P. H. (1988) *J. Mol. Biol. 200*, 709−723.
3. deCrombrugghe, B., Busby, S., and Buc, H. (1984) *Science 244*, 831−838.
4. Ebright, R. H., Ebright, Y. W., and Gunasekera, A. (1989) *Nucleic Acids Res. 17*, 10295−10305.
5. Gaston, K., Chan, B., Lolb, A., Rox, J., and Busby, S. (1988) *Biochem. J. 253*, 809−818.
6. Dalma-Weishausz, D. D., Garternberg, M. R., and Crothers, D. M. (1991) *Nucleic Acids Res. 19*, 611−616.
7. Pyles, E. A., and Lee, J. C. (1996) *Biochemistry 35*, 1162−1172.
8. Takahashi, M., Blazy, B., and Baudeas, A. (1980) *Biochemistry 19*, 5124−5130.
9. Windholz, M., Ed. (1976) in *The Merck Index*, 9th ed., p 353, Merck & Co., Inc., Rahway, NJ.
10. Cerione, R. A., McCarty, R. E., and Hammes, G. G. (1983) *Biochemistry 22*, 769−776.
11. Haughland, R. P. (1996) in *Molecular Probes Handbook of Fluorescent Probes and Research Chemicals* (Larison, K. D., Ed.) 6th ed., p 54, Molecular Probes, Inc., Eugene, OR.
12. Heyduk, T., and Lee, J. C. (1989) *Biochemistry 28*, 6914−6924.
13. Heyduk, T., and Lee, J. C. (1990) *Proc. Natl. Acad. Sci U.S.A. 87*, 1744−1748.
14. Heyduk, T., and Lee, J. C. (1992) *Biochemistry 31*, 5165−5161.
15. Gurney, R. W. (1962) in *Ionic Processes in Solution*, pp 80−112, Dover Publications, Inc., New York.
16. Giraud-Panis, M. J., Toulme, F., Maurizot, J. C., and Culard, F. (1992) *J. Biol. Struct. Dyn. 10*, 295−309.
17. Pyles, E. A., and Lee, J. C. (1998) *Biochemistry 37*, 5201−5210.
18. Shatzky-Schwartz, M., Hiller, A., Reich, Z., Ghirlando, R., Weinberger, S., and Minsky, A. (1992) *Biochemistry 31*, 239−246.
19. Kolb, A., Busby, S., Buch, H., Garges, S., and Adhya, S. (1993) *Annu. Rev. Biochem. 62*, 749−795.

BI972450I