Proposed Structural Mechanism of *Escherichia coli* cAMP Receptor Protein cAMP-Dependent Proteolytic Cleavage Protection and Selective and Nonselective DNA Binding[†,‡]

Sean-Patrick Scott*[,§] and Shadi Jarjous[||]

Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, and School of Medicine, Temple University, Philadelphia, Pennsylvania 19140

Received September 21, 2004; Revised Manuscript Received April 16, 2005

ABSTRACT: The *Escherichia coli* cAMP receptor protein (CRP) displays biphasic characteristics in protease and β-galactosidase induction assays at increasing cAMP concentrations in response to ligand binding at the secondary binding site located between the primary binding site and the DNA binding domain. Two mutants were created to determine the mechanistic reason for the CRP biphasic response by inhibiting binding of cAMP to the secondary site via interference with the Arg 181 interaction with the ligand's phosphate. The S179A/R180D/E181H mutant binds two cAMP molecules per dimer, does not exhibit a biphasic response, lacks selective DNA binding, and has inhibited nonselective DNA binding. The R180K mutant binds four cAMP molecules per dimer, exhibits a biphasic response, nonselective DNA binding similar to CRP, but has inhibited selective DNA binding characteristics. The results are consistent with a 2 × 2-binding site scheme were both primary binding sites must be occupied before the secondary binding sites are occupied. A structural mechanism suggesting the secondary sites are formed by binding of cAMP to the primary sites is proposed. AMMP-generated molecular models suggest that R180 orients E181 to produce selective DNA binding, Arg 169 interactions are necessary for nonselective DNA binding, and the position of Leu 57 inhibits chymotrypsin cleavage of Phe 136. DNA binding results suggest that CRP may be the unknown transcription factor which binds to the temperature sensitive *dsrA* promoter.

The molecule CRP[1] (cAMP receptor protein, also known as catabolite gene activator protein, CAP) binds to the *Escherichia coli lac* promoter in response to increasing amounts of 3′,5′-cyclic adenosine monophosphate (cAMP). CRP has been studied extensively for several decades with recent reviews cataloguing some of this earlier work (*1*−*3*). CRP is a dimer composed of two identical 209-residue monomers (*4*, *5*) containing a well-defined cyclic nucleotide binding domain capable of binding a single anti conformation cAMP molecule (residues 1−129), a hinge region capable of being cleaved by proteases in a ligand-dependent manner (residues 130−139), and a DNA binding domain which recognizes specific DNA sequences (residues 140−209) (*6*). Recent crystal structures revealed that each monomer has a second cAMP binding domain binding syn conformation cAMP between the primary cyclic nucleotide binding domain and the DNA binding domain. Therefore, the CRP dimer is capable of binding a total of four cAMP ligands (*7*, *8*). The presence of cAMP in the secondary binding sites correlates with the length of the DNA fragment used during crystal growth (*8*). A DNA fragment of 38 bp contains cAMP in the secondary sites, while the ligand is absent in crystals grown with a 30 bp DNA fragment. The cocrystals defined the secondary binding site as being formed by residues 56−59, 170, 174, and 177−180 (*9*), with Arg 180 directly interacting with the phosphate of cAMP. The crystals further defined the CRP selective DNA binding to be dependent on residues 180, 181, and 185 (*8*), with residues 180 and 181 occupying the major groove of bound DNA (*10*).

What is not explicit from the crystals is the order in which the sites are filled, although a recent kinetic scheme predicts cAMP binds to the primary binding sites first at micromolar concentrations promoting DNA binding and to the secondary binding sites later, decreasing the affinity of CRP for DNA. The kinetic scheme defined the primary and secondary binding sites as being independent and suggested that the

[†] Supported by American Heart Association Scientist Development Grant 0230378N. S.J. was supported by a grant to Temple University from the Howard Hughes Medical Institute through the Undergraduate Biological Sciences Education Program. Partial salary support for S.-P.S. during production of this paper was provided by the following grants: EY11973 to Stuart E. Dryer and 5R01EY06640-16 to Jacqueline C. Tanaka.

[‡] A preliminary report of this work was presented at the 2003 Biophysical Society Meeting, Abstract 2463.

* To whom correspondence should be addressed. Phone: (713) 743-2693. Fax: (713) 743-2636. E-mail: spscott@uh.edu.

[§] University of Houston.

[||] Temple University.

[1] Abbreviations: CRP, cAMP receptor protein also called catabolite gene-activating protein or CAP; R180K, Lys mutant of Arg 180 in CRP; ADH, triple mutant of Ala 179, Asp 180, and His 181 for Ser 179, Arg 180, and Glu 181 in CRP; cAMP, 3′,5′-cyclic adenosine monophosphate; cGMP, 3′,5′-cyclic guanosine monophosphate; [^{3}H]cAMP and [^{3}H]cGMP, tritium-labeled ligands. PDB-derived models were generated from the PDB structure used as the starting template; the CRP cyclic nucleotide binding site contained in amino acids 1−128 is also known as the high-affinity binding site. The secondary binding site or the low-affinity site is the inhibitory cAMP site located between the primary site and the DNA binding domain. The standard three-letter code for amino acids is used. The amino acid positions in lightface type are in the first of two CRP monomers, and the underlined amino acids are in the second of two CRP monomers.

10.1021/bi0479609 CCC: $30.25
Published on Web 05/27/2005

primary binding sites bind ligand before the secondary binding sites are bound (*11*). The kinetic model is further supported by another recent crystal structure showing three cAMP ligands bound per CRP dimer with cAMP binding to both primary binding sites and a single cAMP in one of the secondary binding sites (personal communication with J. C. Lee). Studies also show that the selective DNA binding occurred at lower concentrations of cAMP and was inhibited at higher ligand concentrations, although nonselective DNA binding was not as greatly inhibited (*12*), suggesting CRP in the presence of low cAMP concentrations is the active form. The DNA binding studies are consistent with the in vivo ligand-dependent biphasic nature seen in *lac* promoter-controlled β-galactosidase assays with increasing concentrations of cAMP (*13*). The active form of CRP can be measured with proteolytic cleavage assays.

Cyclic AMP-dependent CRP chymotrypsin cleavage (*14*) produces a single 12 kDa fragment that is capable of dimerizing and binding cAMP (*15*) and displays biphasic characteristics similar to those exhibited with the β-galactosidase assays and selective DNA binding. At low cAMP concentrations, CRP is cleaved in the hinge region between the primary nucleotide binding domain and the DNA binding domain at Phe 136 (*15*), while cleavage is inhibited at high cAMP concentrations. In light of the cocrystals showing four cAMP ligands bound to the CRP dimer, it can be assumed that the proteolytic inhibition is caused by cAMP binding to the secondary binding sites.

In this report, the secondary binding sites are mutated to decrease the affinity of CRP for cAMP in the secondary binding sites. Molecular models of CRP with cAMP and the CRP antagonist 3′,5′-cyclic guanosine monophosphate (cGMP) bound in the secondary binding sites were compared to direct mutagenesis leading to inhibition of cAMP binding to the secondary sites. cGMP is incapable of activating β-galactosidase transcription (*16*), chymotrypsin cleavage at Phe 136 (*17*), or binding ligand in the secondary binding sites (*18*). The physiological differences between cAMP and cGMP CRP binding suggested that differences in the calculated interaction energies could be used to choose mutations which would inactivate the secondary binding sites.

The molecular model comparisons suggest that the guanosine purine ring may preclude binding of cGMP to the secondary site and that mutations directed against the adenosine purine ring would prove to be fruitless. Therefore, mutations were directed against Arg 180 which interacts with the ribofuranose ring of both ligands to inhibit binding of cAMP to the secondary sites. The R180K conservative mutation was introduced to retain selective DNA binding while inhibiting cAMP binding. The Lys point charge was thought to have to choose between either selective DNA binding or binding cAMP to the secondary site as compared to the "diffuse" Arg charge which is capable of performing both actions simultaneously.

The S179A/R180D/E181H (ADH) mutant was made to ensure that the protein would be incapable of binding the cyclic nucleotide ribofuranose by mutating the positive residue Arg to the negative residue Asp. The additional mutation of the E181 negative charge to the positively charged His was introduced to conserve local and overall charge balance by countering the now negative charge of position 180 and because the size more closely matches that of Glu compared to Lys and Arg. The polar S179 was mutated to Ala to further conserve the local charge environment and allow DH to occupy steric space different from that of RE. The ADH mutant was expected to form the CRP active state in response to binding cAMP in the primary sites and to be able to bind DNA nonselectively as R180 and E181 are likely responsible for only selective DNA binding.

The tritium binding experiments and proteolysis experiments suggest that the ADH mutant binds 2 cAMP molecules per CRP dimer, showing the secondary site cAMP binding is inhibited. Proteolytic cleavage experiments and molecular modeling suggest that the position of Lys 57 is instrumental in the protection of chymotrypsin cleavage of the hinge region at Phe 136 when the proteins adopt the CRP active state. DNA binding experiments in conjunction with molecular modeling suggest that Arg 169 is instrumental in nonselective DNA binding and that Arg 180 positions Glu 181 for selective DNA binding to occur. Finally, the DNA binding results suggest CRP may be the unknown transcription factor which binds to the temperature sensitive *E. coli* promoter *dsrA* (*19*).

METHODS

Static molecular models of ADH, CRP, and R180K are created using the molecular dynamics program AMMP (*20*) with protocols previously described (*21*, *22*). The models can contain two, three, or four molecules of cAMP or four molecules of cGMP, DNA, and RNA polymerase fragments. The X-ray crystal structures have hydrogen atoms and mutations added, while crystallographic waters are removed. The newly added undefined atoms are positioned, while the previously defined atoms are fixed. The atom positions are then refined using all parameters (bonding energies, torsion angles, hybridization energy, nonbonded energies, and bond angles) with conjugate gradient minimization, while the previously defined atoms remain fixed in position. The final minimization step allows all atoms to move without any artificial constraints. The models are minimized until an energy minimum is reached so that the external nonbonded energy of each amino acid could be compared. The static molecular models represent a single possible low-energy allosteric state based on the available CRP crystal structures. The represented modeled allosteric state is only one state in the continuum of allosteric states that together comprise the molecular motions which CRP experiences depending on the presence of ligands, DNA, or RNA polymerase.

Crystal structures 2cgp (*7*), 1g6n (*23*), 1run (*24*), 1lb2 (*25*), 1i5z (personal communication with J. C. Lee and available from the Protein Data Bank), and 1o3q (*8*) are used as starting structures for the molecular models and are chosen as representative CRP structures (see Table 1). Each model is generated from a CRP dimer complex, so those structures with a monomeric subunit in the asymmetric cell have the complement added to form a dimer complex. No other modifications occurred to the template PDB files other than renumbering of the subunits (positions 1−209 for monomer 1 and positions 301−509 for monomer 2, always underlined), the mutation of amino acids (S179 to Ala, R180 to Asp, and E181 to His for ADH and R180 to Lys for R180K), the changing of the ligand from cAMP to cGMP, and/or the removal of the crystallographic waters. Control molecular

Table 1: Crystal Structures Used as Templates for Molecular Modeling[a]

PDB entry	no. of cAMPs/dimer	other	resolution (Å)
1g6n (*23*)	2		2.1
1run (*24*)	2	DNA^1	2.7
$1lb2^M$ (*25*)	2	DNA^1 and polymerase	3.1
1i5z[b]	3		1.9
$1o3q^M$ (*8*)	4	DNA^1	3.0
$2cgp^M$ (*7*)	4	DNA^2	2.2

[a] The superscript M denotes the crystallographic asymmetric unit contains a monomer. The number of cAMPs per dimer is the total number of cAMP molecules bound in the dimer. Other denotes the presence of either of the two DNA fragments (denoted DNA^1 or DNA^2) or RNA polymerase fragments in the crystal structure and retained during molecular modeling. Templates 1o3q and 2cgp are used to compare cAMP and cGMP binding. 1g6n, 1run, 1lb2, 1i5z, and 1o3q are used to compare cAMP binding to ADH, CRP, and R180K. [b] Structure 1i5z was available via personal communication with J. C. Lee and the Protein Data Bank.

modeling experiments (not shown) with 1g6n- and 2cgp-derived CRP with cAMP or cGMP molecular models suggested that the crystallographic waters could be removed with a minimal alteration of results when two static molecular models are compared.

A control model is generated using the X-ray crystal structure with cAMP for comparison with the mutant and cGMP models. The models are visualized for obvious disturbances in the α-carbon backbone trace, for deformations in the cyclic nucleotides and amino acid ring structures, and for obvious differences in the control and mutant model Rhamachandran plots using the program O (*26*). AMMP is used to calculate the nonbonded energy of each amino acid, DNA base, and ligand in each model. The energy differences between corresponding elements in two models are determined by finding the difference between the absolute values of corresponding nonbonded energy and dividing by 2. These values correspond to differences in enthalpy and are color-coded onto an α-carbon trace for easy visualization (see Figures 1 and 4). Water effects are assumed to cancel out due to the "conservative" mutations and overall charge conservation when two models based on the same crystal structure with conservative changes are compared. All energetic differences are therefore assumed to be related to the conservative differences, whether mutations or ligands, in the molecular models.

Control molecular modeling experiments (not shown) suggested comparisons of nonbonded energies with values greater than 10 kcal/mol were significant; 10 kcal/mol became the cutoff value for locating important interactions within the ligand−protein complex. The cutoff value further has the significance of being slightly greater than the estimate of the strongest hydrogen bond, estimated to be 3−9 kcal/mol (12−38 kJ/mol) (*27*). Four other cutoffs were also selected for categorizing the nonbonding energy differences: 1 kcal/mol for random perturbations, 2 kcal/mol for the lower end of estimates for hydrogen bond strength, 5 kcal/mol for the midrange of hydrogen bonding estimates, and >20 kcal/mol or greater than two hydrogen bonds. All estimates were rounded to the nearest whole number before sorting.

CRP mutagenesis is performed using overlap extension PCR (*28*). External primers at the 5′ and 3′ end of CRP are used in conjunction with overlapping internal primers (primers produced by QIAGEN) that contain altered sequences corresponding to the desired S179A, R180D, and E181H mutations in ADH or the R180K mutation in CRP. The clones are restricted with NdeI and EcorI and placed into the pLex vector under the λP_L promoter for overexpression. All clones are sequenced to ensure that the correct mutations were produced.

Overexpression and purification of CRP and the mutants are performed using *E. coli* CA8445-1 (CRP^-, Str^r) with the pRK 248 vector (λcI^{ts}, Tet^r) that encodes a thermolabile λ cI repressor used to control the λP_L promoter in the pLex vector which contains the clone. Modified protocols of Harman and co-workers (*29*) were used for the induction and purification of CRP and mutant proteins. Transformed *E. coli* cells are grown in LB medium with 30 μg/mL tetracycline, 100 μg/mL ampicillin, and 100 μg/mL streptomycin at 30 °C until they reach an absorbance between 0.6 and 0.8 at 600 nm. An equal volume of LB medium heated to 65 °C is then added to the culture, and the culture is allowed to grow for 4 h at 42 °C. The culture is then centrifuged at 3000*g* with a swinging bucket rotor for 15 min at 4 °C, and the bacteria are collected. The bacteria are resuspended in 7.5 mL of working buffer [50 mM $NaPO_4$ (pH 7.8), 1 mM β-mercaptoethanol, 2 mM EDTA, 50 mM NaCl, and 5% (v/v) glycerol] per liter of culture and passed through a French press two times to disrupt the cell membrane and shear the DNA.

The slurry is centrifuged at 15000*g* for 30 min at 4 °C. The resulting supernatant is batch loaded onto DEAE Sephadex A-25 resin (Amersham Biosciences) and the wash through from this column loaded onto Bio-Rex 70 resin (Bio-Rad Laboratories) column. The Bio-Rex 70 column is washed with 5 volumes of buffer and eluted with a salt gradient from 50 to 600 mM NaCl in working buffer. The fractions are analyzed by SDS−PAGE and selected for purity and quantity of the CRP or mutant protein. The desired fractions are pooled and dialyzed against the working buffer. All purifications and dialysis were carried out at 4 °C.

[^{3}H]cAMP and [^{3}H]cGMP equilibrium binding assays are performed at room temperature (20−22 °C) using small Amicon 10 kDa cutoff filtration devices (*30*). Varying amounts of ^{3}H-labeled ligand are mixed with a fixed protein concentration, typically, 80−150 μM dimer, in a total volume of 200 μL containing 50 mM Tris (pH 7.8), 100 mM NaCl, and 1 mM EDTA. Data points with a standard error of the mean (SEM) greater than half the signal were excluded. A linear nonspecific bound (NSB) was subtracted from free ligand concentrations greater than 200 μM. It is difficult to obtain data at the higher concentrations of ligand and, therefore, for the secondary or low-affinity binding sites, due to the increasing NSB and protein concentration limits. The data are normalized for varying protein concentrations and then fit with eq 1 or 2 in Tablecurve 2D from Jendel Scientific with SEM used to weight the data.

Cyclic AMP binding by ADH and all cGMP binding are fit with a two-binding site model (Scheme 1 and eq 1). Primary binding sites 1 and 2 are initially equivalent in the two-binding site model with the dimer composed of two equivalent monomers on the basis of the assumption. The two-binding site model kinetic scheme and equation are

Scheme 1

$$[P_f] + [cN_f] \overset{K_1}{\longleftrightarrow} [P{\cdot}cN] + [cN_f] \overset{K_2}{\longleftrightarrow} [P{\cdot}cN^2] + [cN_f]$$

$$\frac{[\text{bound cNMP}]}{[\text{dimer}]} = \frac{2K_1cN_f + 2K_1K_2cN_f^2}{1 + 2K_1cN_f + K_1K_2cN_f^2} \quad (1)$$

where $[P_f]$ is the free protein dimer concentration, N_f is the free ligand concentration, P is the dimer with one or two bound ligands, and K_1 and K_2 are the equilibrium association constants. The factors in the numerator and denominator before the K's in the equation are statistical factors used to balance the number of available binding sites and bound ligands. The 2 preceding K_1 accounts for the two free binding sites capable of binding the ligand. The 2 before the numerator K_1K_2 accounts for two ligands being bound to the protein and is missing from the denominator because the dimer concentration is considered here. The division of each term by $[P_f]$ produces the 1 term. The numerator is consciously not divided by 2 to make it equivalent to a two-site Adair equation (1,1) to show the maximum number of bound ligands on the plot.

The cAMP data with R180K and CRP are fit with a 2 × 2-site binding equation. The first two binding sites are equivalent to binding sites 1 and 2 and represent the primary binding sites. Secondary binding sites 3 and 4 bind ligand after primary binding site 2 is occupied. The absence of X-ray data with ligands bound to binding sites 1 and 3 without binding site 2 also containing ligand supports this assumption. Binding sites 3 and 4 are initially equivalent, but are differentiated when binding site 3 binds ligand. The assumption is made even though the CRP X-ray crystal structure suggests structural differences in the binding sites (*6*). The kinetic scheme [similar to that proposed by the Lee group (*11*)] and equation are

Scheme 2

$$[P_f] + [cN_f] \overset{K_1}{\longleftrightarrow} [P{\cdot}cN_1] + [cN_f] \overset{K_2}{\longleftrightarrow} [P{\cdot}cN_1^2] + [cN_f] \overset{K_3}{\longleftrightarrow} [P{\cdot}cN_1^2{\cdot}cN_2] + [cN_f] \overset{K_4}{\longleftrightarrow} [P{\cdot}cN_1^2{\cdot}cN_2^2] + [cN_f]$$

$$\frac{[\text{bound cNMP}]}{[\text{dimer}]} = \frac{2K_1cN_f + 2K_1K_2cN_f^2 + 6K_1K_2K_3cN_f^3 + 4K_1K_2K_3K_4cN_f^4}{1 + 2K_1cN_f + K_1K_2cN_f^2 + 2K_1K_2K_3cN_f^3 + K_1K_2K_3K_4cN_f^4} \quad (2)$$

where $[P_f]$ is the free protein dimer, [P] is the concentration of bound dimer in each state, N_f is the free ligand concentration, cN_1 and cN_2 are the primary and secondary binding sites, respectively, and K_1, K_2, K_3, and K_4 are the association constants for ligands 1, 2, 3, and 4, respectively. The numbers in each term are statistical weights accounting for the number of binding sites available and the number of ligands bound.

In the numerator, the K_1 and K_1K_2 coefficients are equivalent to those in the two-binding site model. The three bound ligands and the two equivalent binding sites generate the 6. The four bound ligands account for the 4. In the denominator, the 2 before $K_1K_2K_3$ denotes the number of free binding sites, whereas there is no factor before $K_1K_2K_3K_4$ as this is in terms of protein concentration in the denominator. The 1 term derives from dividing each term by $[P_f]$. To compare to a four-site Adair equation, the numerator needs all terms divided by 4, giving 0.5, 0.5, 1.5, and 1. The four-binding site Adair equation coefficients (numerator 1, 3, 3, 1 and denominator 1, 4, 6, 4, 1) are not equal to the 2 × 2-binding site model due to the 2 × 2-site model assumption that the four binding sites are not initially equivalent. A 2 × 1-site model is a variation of the 2 × 2-site model described above. The assumption of the negative cooperativity being so immense that a fourth cAMP molecule is unable to bind as suggested by the 1i5z crystal structure differentiates this model from the 2 × 2-site model. The fourth term in the kinetic scheme is dropped, and the terms including K_4 in the numerator and denominator are removed. There is no change in the coefficients as the total number of binding domains remains the same.

Chymotrypsin proteolytic assays are performed with no ligand or varying cAMP concentrations at 37 °C with concentrations of protein and α-chymotrypsin (Sigma-Aldrich Co., catalog no. C4129, EC 3.4.21.1) of 20 μM and 12 μg/mL (~0.5 μM), respectively, in reaction buffer [50 mM sodium phosphate (pH 7.8), 100 mM NaCl, and 1 mM EDTA]. The reaction progresses for 15 min when the cAMP concentration is varied and from 0.5 to 300 min during time course experiments. The reaction is stopped by adding 25 μL of 2× gel loading dye [100 mM Tris-HCl (pH 6.8), 200 mM dithiothreitol, 4% (w/v) SDS, 0.2% bromophenol blue, 20% glycerol, 0.5 μg of BSA, and 2 mM PMSF] to 25 μL of sample followed by heating for at least 1 min at >80 °C; 25 μL of the sample is electrophoresed on 13 cm × 14 cm × 0.1 cm 15% acrylamide gels. The gels are digitized and analyzed with Scion Imaging software by Scion Corp.

The CRP cleavage rate is determined for several ligand concentrations at fixed protein and chymotrypsin concentrations, similar to what had been previously done with subtilisin (*31*). A time course experiment produces a decrease in the amount of uncleaved protein (U) and a proportional increase in the amount of cleaved protein (C). The cleavage rate of U and the formation rate of C are equivalent under the conditions that were used, although C is further degraded during longer cleavage times (>90 min) to fragments not visible on a gel. The fixed protein and chymotrypsin concentrations, as well as the detection limits of the gels, ensure that the visible reaction took place under steady-state conditions.

Taking the log of the amount of uncleaved protein produces a linear transform when plotted against reaction time as a further indication that the reaction occurred at steady state. The slope of the line was taken as the cleavage rate. A nonlinear fit is applied to the cleaved protein so that the fit is not biased by an inaccurate determination of the final amount of protein. The reaction rates were left in terms of arbitrary units per minute (AU/min) as the amount of protein is determined in Scion with AU.

Cleavage and formation reaction rates are determined from gels with at least five reliable data points in the 10−90% amount of protein range. Extremely slow or fast rates sometimes could only have their formation or cleavage rate

determined. The two rates were averaged when both were available and from the same gel, and each rate was calculated with an average of 8.5 points in the 10−90% range. In addition, the formation rates have at least three data points defining the plateau for the fit.

The kinetic scheme below is assumed

Scheme 3

$$\begin{array}{ccccccc} & & K_1 & & K_3 & & \\ [cAMP_{Free}] + [CRP_U] & \Leftrightarrow & [cAMP_{Free}] + [CRP_U{\bullet}cAMP] & \Leftrightarrow & [CRP_U{\bullet}cAMP^2] + [cAMP_{Free}] & & \\ \downarrow v_0 & & \downarrow v_1 & & \downarrow v_2 & & \\ [cAMP_{Free}] + [CRP_C] & \Leftrightarrow & [cAMP_{Free}] + [CRP_C{\bullet}cAMP] & = & [CRP_C{\bullet}cAMP] + [cAMP_{Free}] & & \\ & & K_C & & & & \end{array}$$

where CRP_U is the uncleaved monomer capable of binding two ligands, one in the primary binding site and one in the secondary binding site, CRP_C is the cleaved monomer which contains only the primary binding site (*15*, *32*), and v_0, v_1, and v_2 are the irreversible cleavage rates when the monomer binds zero, one, and two ligands, respectively. The individual rates combine to produce the total cleavage rate (v_{total}) at any given ligand concentration. The reduced equation for describing the relationship for monomeric CRP is

$$v_{total} = \frac{v_{no\ lig} + v_{SatADH}K_1[cAMP] + v_{SatCRP+R180K}K_1K_3[cAMP]^2}{1 + K_1[cAMP] + K_1K_3[cAMP]^2} \quad (3)$$

where $v_{no\ ligand}$, v_{SatADH}, and $v_{SatCRP+R180K}$ correspond to v_0, v_1, and v_2, respectively. These rates are determined at limiting conditions of no ligand or saturating ligand and assumed to be the same for all three proteins. K_1 and K_3 are the equilibrium constants for cAMP binding to the primary and secondary binding domains, respectively. K_3 is used instead of K_2 to keep the nomenclature between monomeric CRP and dimeric CRP equivalent. K_C is the constant for binding of ligand to the cleaved protein and is not accounted for as the concentration of cAMP is much greater than the total amount of CRP such that the effect should be negligible. This is not an Adair type equation as previously used (*31*) since the ligand binding is assumed to be sequential with the primary binding site needing to bind ligand and forming the secondary binding site which can then bind ligand.

This equation is reduced for ADH to

$$v_{total} = \frac{v_{no\ lig} + v_{SatADH}K_1[cAMP]}{1 + K_1[cAMP]} \quad (4)$$

because the secondary binding site is not capable of binding cAMP. Equation 4 is expanded for fitting to a dimer instead of a monomer:

$$v_{total} = \frac{v_{no\ lig} + v_{SatADH}K_1[cAMP] + 2v_{SatADH}K_1K_2[cAMP]^2}{1 + K_1[cAMP] + v_{SatADH}K_1K_2[cAMP]^2} \quad (5)$$

where K_2 is the second primary binding site association constant. The coefficient of 2 is added to account for the two monomers that are assumed to be cleaved at the same reaction rate. The v_{total} values at various cAMP concentrations with ADH, CRP, and R180K were fit with Tablecurve 2D using user-defined eqs 3−5 to determine K_1 and v_{SatADH} for monomeric ADH, K_1 and K_3 for CRP and R180K, and K_1, K_2, and v_{SatADH} for dimeric ADH.

DNA binding assays are used to determine the ability of the ADH and R180K to selectively and nonselectively bind DNA. PCR methods are used to generate DNA fragments of T7 polymerase (2682 bp) (*33*), LeuO protein (1070 bp) (*34*), the thermolabile cIλ repressor (759 bp) (*35*), a fragment of CRP (496 bp), the *lac* promoter (345 bp), the *dsrA* promoter (209 bp) (*19*), and the *mcl* promoter (95 bp) (*36*) for the DNA binding experiments. Purified DNA from Bl-21 DE3 *E. coli* (Novagen Inc.) is the DNA template for T7 polymerase, LeuO, the *dsrA* promoter, and the *mcl* promoter. The *lac* promoter template is generated from the pAlter-1 vector (Promega Corp.) and the thermolabile λ repressor template from the pRK 248 vector (*37*). The *lac* promoter contains two CRP binding sites.

The DNA binding reactions for the gel in Figure 6A use all the DNA fragments but that derived from CRP, and reactions for panels B−D of Figure 6 include no *lac* promoter DNA. DNA binding followed the method described by Crothers and co-workers (*38*) and described previously (*30*) with minor modifications. Binding reaction mixtures are incubated at room temperature in the presence of 200 μM cAMP (Figure 6A) or 2 mM cAMP (Figure 6B−D) and run on 9 cm × 8 cm × 0.1 cm gels composed of 6% (w/v) total acrylamide (29:1 acrylamide:bisacrylamide). Final solution conditions in the binding reactions are 30 mM sodium acetate, 10 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 10 mM $MgCl_2$, 1 mM DTT, and 10% glycerol. Reaction mixtures have a total volume of 25 μL containing ~100 ng of each DNA fragment and the indicated concentration of ADH, CRP, or R180K. Gels are stained by being soaked in ethidium bromide and photographed under UV illumination.

RESULTS

Initial molecular modeling attempts using 1o3q- and 2cgp-derived molecular models with cAMP and cGMP are averaged and compared to suggest mutations for inactivation of the secondary cyclic nucleotide binding domain (Figure 1). The difference in the nonbonded external energy for amino acids in the cGMP model and the cAMP model is shown in the α-carbon trace in Figure 1A, with the amino acid locations or ligand (space filling) color-coded for their differences divided into five ranges: blue for 0−1 kJ/mol, cyan for 2−4 kJ/mol, violet for 5−9 kJ/mol, yellow for 10−19 kJ/mol, and red for >20 kJ/mol. The residues which have large nonbonded external energy differences (yellow and red residues) are spread throughout the dimer with Glu 34 and Lys 35 in β-strand 3, Glu 54, Glu 55, and Lys 57 in the β4−β5 loop, and Asp 155 in the αD−β9 loop showing differences in both monomers. Residues Glu 37 in β-strand 3, Asp 53 in the β4−β5 loop, Arg 82 and Thr 89 in β-strand 7, Tyr 100 and Lys 101 in α-helix B, Glu 129 in α-helix C, Asp 138 in the αC−αD loop, Arg 142 in α-helix D, and Arg 185 in α-helix F exhibit differences in one of the two monomers.

Lys 57 is the only residue which faces the purine ring and experiences a calculated energy difference when cGMP

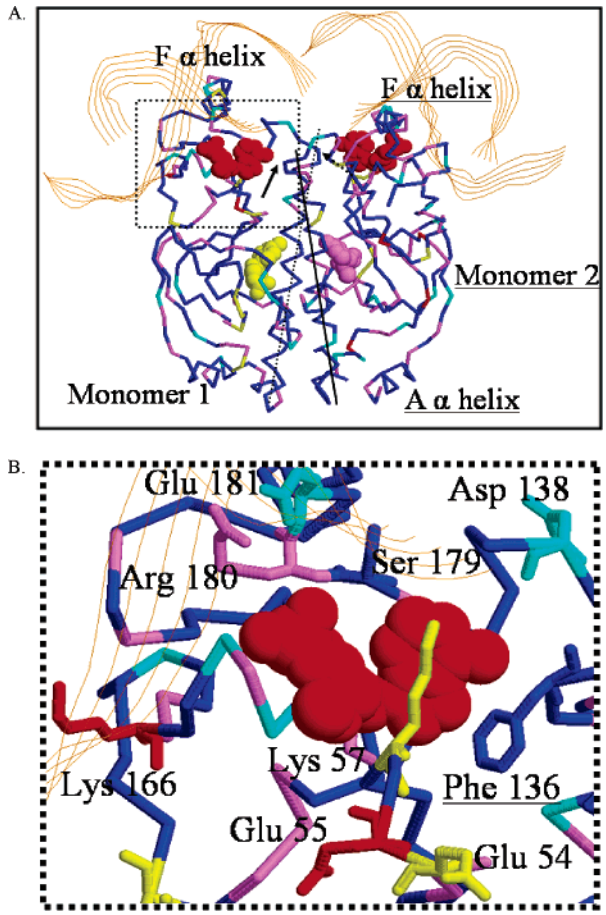

FIGURE 1: Molecular modeling suggests syn cGMP purine ring differences inhibit binding to the CRP secondary binding sites. (A) Dimer 1o3q- and 2cgp-derived CRP molecular models with cAMP or cGMP had the external nonbonded energy for each amino acid determined by AMMP. The average absolute ($cGMP_{average}$ − $cAMP_{average}$)/2 external nonbonded energy difference is shown on the 2cgp crystal structure α-carbon trace. The amino acid locations or ligands (space filling) are color-coded for their differences: blue for 0−1 kJ/mol, cyan for 2−4 kJ/mol, violet for 5−9 kJ/mol, yellow for 10−19 kJ/mol, and red for >20 kJ/mol. The bound DNA is colored orange. The labels are locations in either monomer 1 or monomer 2 (underlined). The two lines are the path of the C α-helices for monomer 1 (dotted) and monomer 2 (solid). The arrows point to the position of Phe 136 for monomer 1 (dotted tail) and monomer 2 (solid tail). The dotted box contains a secondary cAMP binding site expanded in the next panel. (B) Close-up of a secondary binding domain. Glu 54, Glu 55, and Lys 57 are on the β4−β5 loop. Ser 179, Arg 180, and Glu 181 are located at the beginning of α-helix F. Asp 138 is located in the αC−αD loop. Lys 166 is between β-strand 10 and α-helix E. Phe 136 is located on α-helix C of the second CRP monomer. All labeled residues except Glu 54, Glu 55, Asp 138, and Lys 166 face the bound cAMP ligand.

replaces cAMP in the models (Figure 1B). It lies across the purine ring interacting with the bound DNA and comprises part of β-strand 5 that is important for CRP activation. Also, Lys 57 interacts with the electron rich purine and bound DNA and therefore should not select cAMP over cGMP. Further, the models seem to indicate that there are differences in the cAMP and cGMP purine ring structure that may inhibit binding of cGMP to the secondary binding domain and not protein interactions with ligand. For these three reasons, Lys 57 is excluded as a candidate, leaving no viable candidates for inhibition of the binding of cAMP to the secondary cyclic nucleotide binding domain. Residues which interact with the ribofuranose were next targeted to remove cyclic nucleotide binding. Arg 180 interacts with the phosphate of the ribofuranose ring and is the best candidate for mutation. Ser 179 also interacts with the ribofuranose ring and may be a determinant for syn and anti purine ring conformation so was also chosen as a candidate. It should be noted that the Ser-Arg sequence is the reverse of an important Arg-Ser (or the similar Arg-Thr) sequence in the family of cyclic nucleotide binding domains represented by the CRP primary cyclic nucleotide binding domain. Glu 181 is also a candidate for mutation, but mutation would likely result in a loss of DNA specificity. The two mutants on which we decided were the Ser179Ala/Arg180Asp/Glu181His (ADH) and Arg180Lys (R180K) mutants. The ADH mutant was made to completely inhibit binding of cAMP to the secondary cyclic nucleotide binding domain, and R180K was thought to be able to move the positive charge far enough from the secondary binding site to result in weakened cAMP binding while still being able to retain the specific DNA binding capabilities of CRP.

Purification and expression of the mutants showed qualitative differences between their purification and expression characteristics and those of CRP. First, empirically, ADH purifications produced more protein per liter of growth medium than R180K which produced more than CRP. Second, purification of the proteins was to have a two-column system: A25 anion-exchange resin used to remove contaminating proteins followed by BioRex 70 cation-exchange resin that would retain the protein until a salt ramp was used to elute the protein from the BioRex 70 resin. The BioRex 70 column did not readily retain ADH, although it did retard the progress until late in the washing step when contaminating proteins were nonexistent. Third, empirically, ADH and R180K seemed to be less stable to temperature changes than CRP. Finally, only CRP was able to elicit positive results when using MacKonkey Agar (results not shown) to show β-galactosidase activity in response to overexpression of CRP in the presence of 10 mM cAMP.

Overloaded protein analyzed by SDS−PAGE was purified to >95% for ADH, CRP, and R180K. The concentrations of the protein and the results were standardized for all the following experiments using an extinction coefficient of 20 400 M^{-1} cm^{-1} at 278 nM (*39*). Many of the experiments were carried out in tandem with ADH, CRP, and R180K to minimize differences caused by reagents and other unforeseen variations in procedure. CRP could be concentrated to 22.5 mg/mL, while R180K and ADH were limited to concentrations no greater than 5.0 and 7.5 mg/mL, respectively, using Amicon Bioseparation's Centriprep and Centricon centrifugal filter devices.

[^{3}H]cAMP and [^{3}H]cGMP binding experiments are consistent with CRP and R180K binding four cAMP ligands at saturating cAMP concentrations and two ligands at saturating cGMP concentrations (Figure 2 and Table 2). ADH binds two ligands at saturating concentrations of both cAMP and cGMP. The two-binding site model is used to fit the cGMP curves (r^2 = 0.96 for ADH, r^2 = 0.87 for CRP, and r^2 = 0.95 for R180K) as well as the ADH cAMP binding curve (r^2 = 0.89). All cGMP fits had random residuals (not shown) throughout the range of ligand concentrations that were tested using the two-site model. The 2 × 2-binding site model was used to fit the CRP (r^2 = 0.91) and R180K (r^2 = 0.97) cAMP data. The ADH cAMP data 2 × 2-site fit (r^2 = 0.83, not

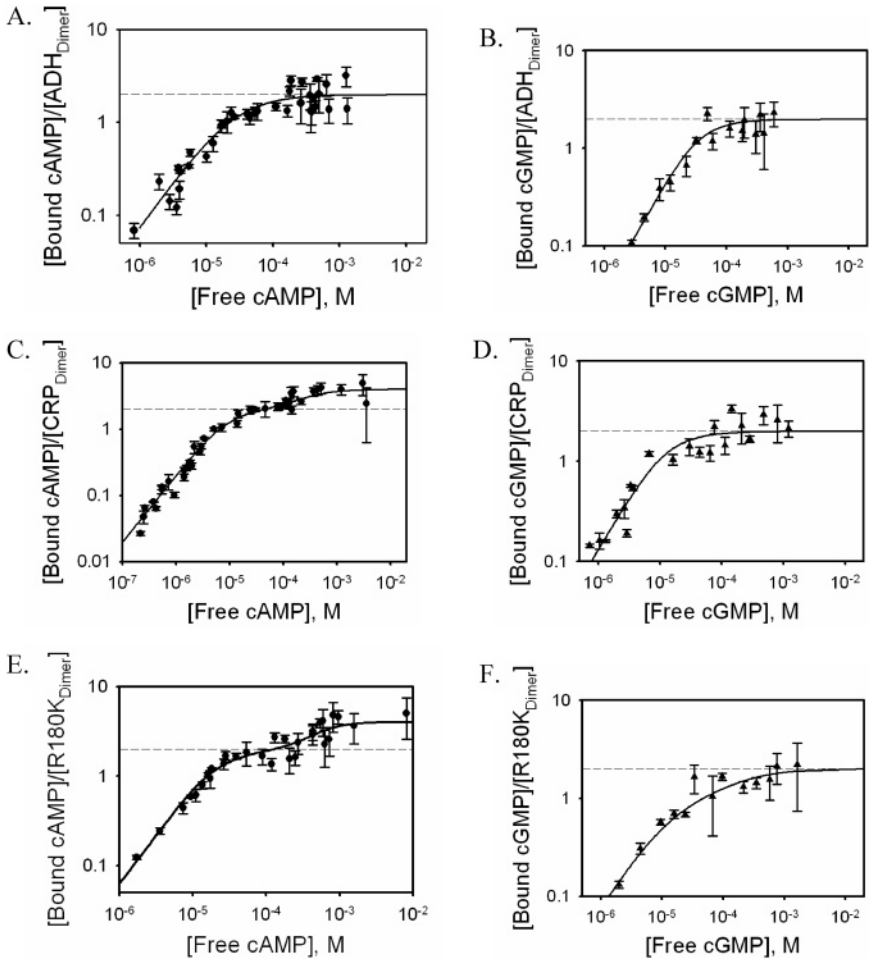

FIGURE 2: ADH dimer which binds only two ligands at saturating cAMP concentrations. Data for binding of [^{3}H]cAMP−ADH (A), [^{3}H]cGMP−ADH (B), R180K (D), and CRP (F) are fit with the two-site binding model (eq 1). Data for binding of [^{3}H]cAMP−CRP (C) and R180K (E) are fit with the 2 × 2-site binding model (eq 2). The dotted line denotes two ligands per dimer. The error bars are in terms of the standard error of the mean. The solid line is the fit. The association constants are listed in Table 2.

Table 2: [^{3}H]cAMP and [^{3}H]cGMP Binding Constants for CRP and the ADH and R180K Mutants[a]

	K_{a1} (M^{-1})	K_{a2} (M^{-1})	K_{a3} (M^{-1})	K_{a4} (M^{-1})	r^2	no. of points[b]
cAMP						
ADH^2	37000	52500			0.89	37
CRP$^{2\times2}$	98200	227900	2000	7200	0.91	40
R180K$^{2\times2}$	29700	133200	30	144900	0.97	30
cGMP						
ADH^2	17400	82900			0.96	15
CRP^2	62800	182300			0.87	20
R180K^2	40400	11000			0.95	14

[a] The weighted data were fit with a two-binding site model (superscript 2) or a 2 × 2-site model (superscript 2×2). r^2 denotes the goodness of fit. [b] Number of data points used in the fit.

shown) consistently overestimated the number of ligands bound per dimer at high cAMP concentrations compared to the random residuals of the two-site model fit. The CRP and R180K cAMP data two-site fit (r^2 = 0.89 and 0.96, respectively, not shown) consistently underestimated the number of bound ligands per dimer at high cAMP concentrations compared to random residuals with the 2 × 2-site model fits. The CRP and R180K cAMP data were also fit to a 2 × 1-site model producing the same r^2 and K_1's and K_2's similar to those of the 2 × 2-site model with larger K_3's accounting for the loss of a site (K_{3CRP} = 5100 M^{-1} and K_{3R180K} = 1700 M^{-1}). The residuals showed the fit consistently underestimated the number of molecules bound as occurred with the two-site model fitting.

Each data set represents several experiments and purifications normalized to account for varying protein concentrations and specific activity used for each experiment with an average of 35 data points used for cAMP fits and 16 data points for cGMP fits. All the fits showed an increased affinity for binding the second cyclic nucleotide in the dimer except for R180K with [^{3}H]cGMP. The secondary site had less affinity for cAMP than the primary site, and ADH exhibited no binding to the secondary sites. The differences in the secondary binding site cAMP affinity by CRP and R180K could be due to the quality of data at the higher cAMP concentrations.

The association constants were converted to ΔG, and the differences among CRP, ADH, and R10K (Table 3) ligand binding were compared. All three proteins show a positive cooperativity (average of 2.2 kJ/mol) in the primary binding sites with [^{3}H]cAMP with ADH (−53.0 kJ/mol) and R180K (−54.7 kJ/mol) having nearly equivalent total binding energies. CRP (−59.0 kJ/mol) produces slightly more energy when binding [^{3}H]cAMP to the primary binding sites. CRP (−40.9 kJ/mol) produces more energy than R180K (−37.9 kJ/mol) when binding ligands to the secondary binding sites,

Table 3: Association Constants Converted to ΔG Values[a]

cAMP	ΔG_1	ΔG_2	ΔG_3	ΔG_4	ΔG_{tot}	ΔG_{12}	ΔG_{34}	ΔG_{tot1+2}	ΔG_{tot3+4}
ADH[2]	−26.1	−27.0			−53.0	0.9		−53.0	
CRP[2×2]	−28.5	−30.5	−18.9	−22.2	−99.9	2.1	3.1	−59.0	−40.9
R180K[2 ×2]	−25.5	−29.2	−8.4	−29.4	−92.6	3.7	21.0	−54.7	−37.9
cGMP									
ADH[2]	−24.2	−28.1			−52.3	3.9		−52.3	NA
CRP[2]	−27.4	−30.0			−57.4	2.6		−57.4	NA
R180K[2]	−26.3	−23.1			−49.3	−3.2		−49.3	NA

[a] ΔG_1, ΔG_2, ΔG_3, and ΔG_4 are the binding energies for K_{a1}, K_{a2}, K_{a3}, and K_{a4}, respectively, converted using the relation $\Delta G = -RT \ln K_a$, where $T = 298$ K and $R = 8.314$ J mol^{-1} K^{-1}. ΔG_{12} and ΔG_{34} are the cooperativity energies in the primary and secondary binding sites, respectively, calculated by the formula $\Delta G_{12} = 2\Delta G_1 - \Delta G_1 - \Delta G_2$. ΔG_{tot1+2} and ΔG_{tot3+4} are the amounts of energy contributed from binding the two ligands in the primary and secondary sites, respectively, and ΔG_{tot} is the total energy contributed by all ligands. All entries are in units of kilojoules per mole.

Table 4: ADH, R180K, and CRP cAMP Binding Constants Derived from Proteolytic Cleavage Reactions[a]

	K_{a1}	K_{a2}	K_{a3}	r^2	ΔG_1	ΔG_2	ΔG_3	$\Delta G_{1prot-trit}$	$\Delta G_{2prot-trit}$	$\Delta G_{3prot-trit}$	v_1
ADH[M]	13700			0.913	−23.5			2.5			0.251
ADH[D]	35000	5600		0.914	−25.9	−21.4		0.2	5.5		0.126
CRP	8300		700		−22.4		−16.1	6.1		2.8	0.251[F]
R180K	6100		1300		−21.6		−17.9	3.9		−9.4	0.251[F]

[a] The ADH, CRP, or R180K ν_{tot} curves were fit with single-variable rate monomeric monophasic model (ADH[M]), a single-variable rate dimeric monophasic model (ADH[D]), or a fixed rate monomeric biphasic model (v^F). cAMP association constants K_{a1}, K_{a2}, and K_{a3} are for binding sites 1, 2, and 3, respectively. ΔG_1, ΔG_2, and ΔG_3 are the calculated binding energies. $\Delta G_{1prot-trit}$, $\Delta G_{2prot-trit}$, and $\Delta G_{3prot-trit}$ are the differences between the ΔG_1, ΔG_2, and ΔG_3 binding energies determined using [^{3}H]cAMP binding and protease cleavage reactions. All ΔG values are in units of kilojoules per mole.

and both exhibit positive cooperativity (3.1 and 21.0 kJ/mol, respectively). ADH, CRP, and R180K bind cGMP with an average of −53.0 kJ/mol. ADH and CRP display positive cooperativity (3.9 and 2.6 kJ/mol, respectively), while R180K has negative cooperativity (−3.2 kJ/mol).

The energies are used only for comparative purposes, and the small differences and quality of data suggest that the binding constants might be more similar than those calculated from the data. The positive cooperativity trends are likely real, although the negative cooperativity of R180K with cGMP suggests caution when comparing the values. The quality of data did not allow exclusion of a four-site Adair model which can fit the data with the equivalent r^2 as the 2 × 2-site model producing binding energies distributed more evenly over the four-site Adair model's four "equivalent" cyclic nucleotide binding sites. The binding data clearly suggest the ADH mutation has inhibited binding of cAMP to the secondary cyclic nucleotide binding domain.

Chymotrypsin proteolytic cleavage experiments (Figure 3) are performed to determine whether the biphasic proteolytic cleavage pattern is due to binding of cAMP to the secondary binding sites and to clarify the tritium binding experiments. The first set of ADH proteolytic cleavage experiments (Figure 3A) suggests that the biphasic cleavage pattern seen with CRP and R180K (Figure 3B,C) is due to binding of cAMP in the secondary nucleotide binding domains and not due to binding of cAMP in the second of the primary binding sites. Further, the data show that ADH is not capable of binding cAMP in the secondary binding sites and that the two-binding site model was the correct model for determining cAMP tritium binding constants. The biphasic pattern for CRP and R180K suggests a four-binding site Adair model is inappropriate for determining cAMP binding constants. Finally, chymotrypsin was unable to cleave ADH, CRP, or R180K in the presence of varying concentrations of cGMP (not shown).

ADH reaction rates are fit with monophasic equations (Figure 3D) with ADH treated as a monomer or a dimer with each monomer binding a single ligand (v_1) (Table 3). CRP and R180K are fit with biphasic rate equations (panels E and F of Figure 3, respectively) that contain fixed reaction rates. The data are not sufficiently complete to fit the CRP and R180K reaction curves with a 2 × 2-binding site dimeric equation. It was assumed that the ADH, CRP, and R180K reaction rates for zero, one, or two bound ligands per monomer would not vary due to the placement of the mutations outside the peptide region that chymotrypsin binds and cleavage occurs at Phe 136.

The average reaction rates used for fitting the data for zero and two bound ligands were determined at the limits of the CRP and R180K curves (0 mM cAMP for zero bound cAMP molecules and 40 mM cAMP for two bound cAMP molecules) to be 0.0045 and 0.0073 AU/min, respectively. The reaction rate with one ligand bound is calculated when the ADH curve is fit to the monophasic equation, determined to be 0.251 AU/min, and fixed in the biphasic rate equation for use with CRP and R180K. The ADH rate of 0.251 is near the average rate of 0.253 for the three proteins and is used here as it is possible to saturate the single primary site of each monomer. The rates for variable one-ligand bound fits of CRP and R180K (not shown) are 0.261 and 0.248, respectively, with the r^2 values being the same as those for the fixed rates. Phe 136 is 34 times more likely to be exposed to solvent and cleaved by chymotrypsin when a single ligand of cAMP is bound to the CRP monomer than when two ligands are bound and is more than 50 times more likely to be cleaved than in the absence of ligand according to these data. The rate decreased to nearly half (0.126 AU/min) with ADH determined by a dimeric equation.

The ADH, CRP, and R180K cAMP association constants are also determined during fitting of the total cleavage rate

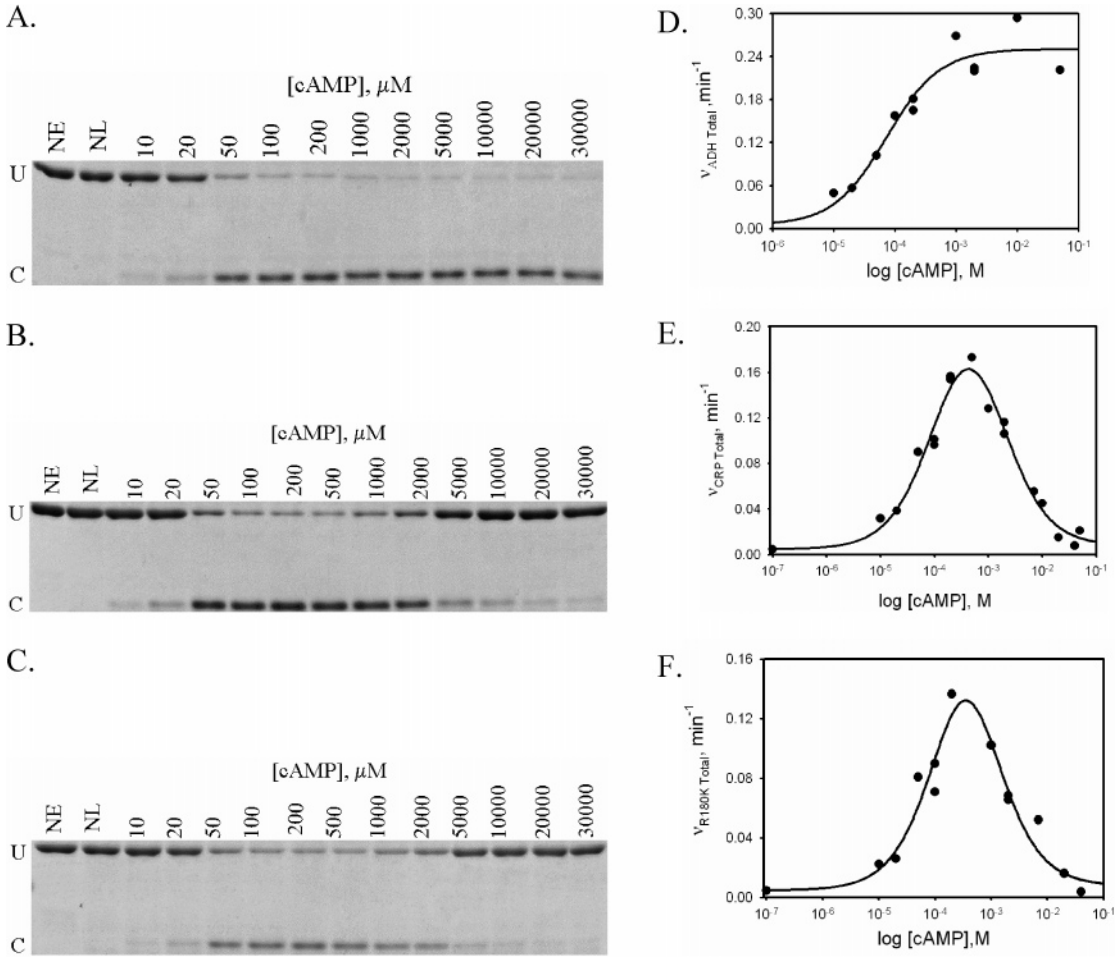

FIGURE 3: ADH is not protected from chymotrypsin at saturating concentrations of cAMP. ADH (A) has a monophasic cAMP-dependent response to chymotrypsin cleavage, and CRP (B) and R180K (C) have biphasic cAMP-dependent responses to chymotrypsin cleavage. U denotes uncleaved protein, and C denotes chymotrypsin-cleaved protein. NE lanes had no protease added, and NL lanes had protease but no ligand. The plots of measured v_{total} vs log[cAMP] for ADH (D), CRP (E), and R180K (F) were fit (solid line) by a monophasic rate equation ($v_{ADHtotal}$, eq 4 or 5) or a biphasic rate equation ($v_{CRPtotal}$ and $v_{R180Ktotal}$, eq 3). The associated v, K_a, and r^2 values for each fit are listed in Table 4.

cAMP concentration dependence (Table 3). The monomeric fits assume the association constants are for binding of the ligand to the primary or secondary binding sites, and two chymotrypsin proteins can bind the protein dimer at the same time. The association constant for the primary binding site shows less affinity compared to that found using the tritium binding method ($K_{a1H}{}^3 = 98\ 200\ M^{-1}$ and $K_{a1proteolysis} = 8300\ M^{-1}$ for CRP). Binding of a ligand to the secondary binding sites precludes chymotrypsin cleavage and therefore allows a more accurate determination of K_{a3} in the proteolytic assay. The pattern of primary binding site affinity for cAMP (ADH > CRP > R180K) is not the same as in the tritium binding experiments (CRP > ADH > R180K), but this could be due to the quality of tritium binding experiments at high ligand concentrations affecting the calculated K_a's for both the primary and secondary binding sites. The affinities of the CRP ($K_{a3H}{}^3 = 2000\ M^{-1}$ and $K_{a3proteolysis} = 700\ M^{-1}$) and R180K ($K_{a3H}{}^3 = 30\ M^{-1}$ and $K_{a3proteolysis} = 1300\ M^{-1}$) secondary binding sites are likely more similar than the tritium experiments suggested. The $K_{a1proteolysis}$ of ADH ($35\ 000\ M^{-1}$) is nearly equal to $K_{a1H}{}^3$ ($37\ 000\ M^{-1}$) when the monomer restriction is removed, although $K_{a2proteolysis}$ ($5600\ M^{-1}$) is 1 order of magnitude smaller than $K_{a2H}{}^3$ ($52\ 500\ M^{-1}$).

The cAMP proteolytic association constants were also converted to binding energies (ΔG) for comparison with tritium binding energies. The near-equivalent amount of binding energy for the first ligand binding ADH in the dimeric model ($\Delta G_{1H}{}^3 = -26.1$ kJ/mol and $\Delta G_{1proteolysis} = -25.9$ kJ/mol) suggests the decrease in the K_a's when comparing the tritium and proteolytic cleavage binding assays is due to the assumption that two chymotrypsin molecules can bind the ADH dimer at the same time. The inability of two chymotrypsins to bind to the ADH dimer reduces the apparent binding energy of the second binding step and effectively decreases the accompanying K_{a2} in the dimeric models. The decrease in monomeric model $\Delta G_{1proteolysis}$ (2.5 kJ/mol) is nearly half the loss of the dimeric model $\Delta G_{2proteolysis}$ (5.5 kJ/mol) because the reduction in energy is mathematically distributed across the binding reactions in monomeric models. The binding energy difference (2.8 kJ/mol) for CRP's secondary binding site is attributed to having a smaller energetic penalty for binding a second chymotrypsin simultaneously, which is necessary with the primary binding sites. The penalty for not being able to bind two chymotrypsins is likely near the minimal amount of 2.5−2.8 kcal/mol per monomer.

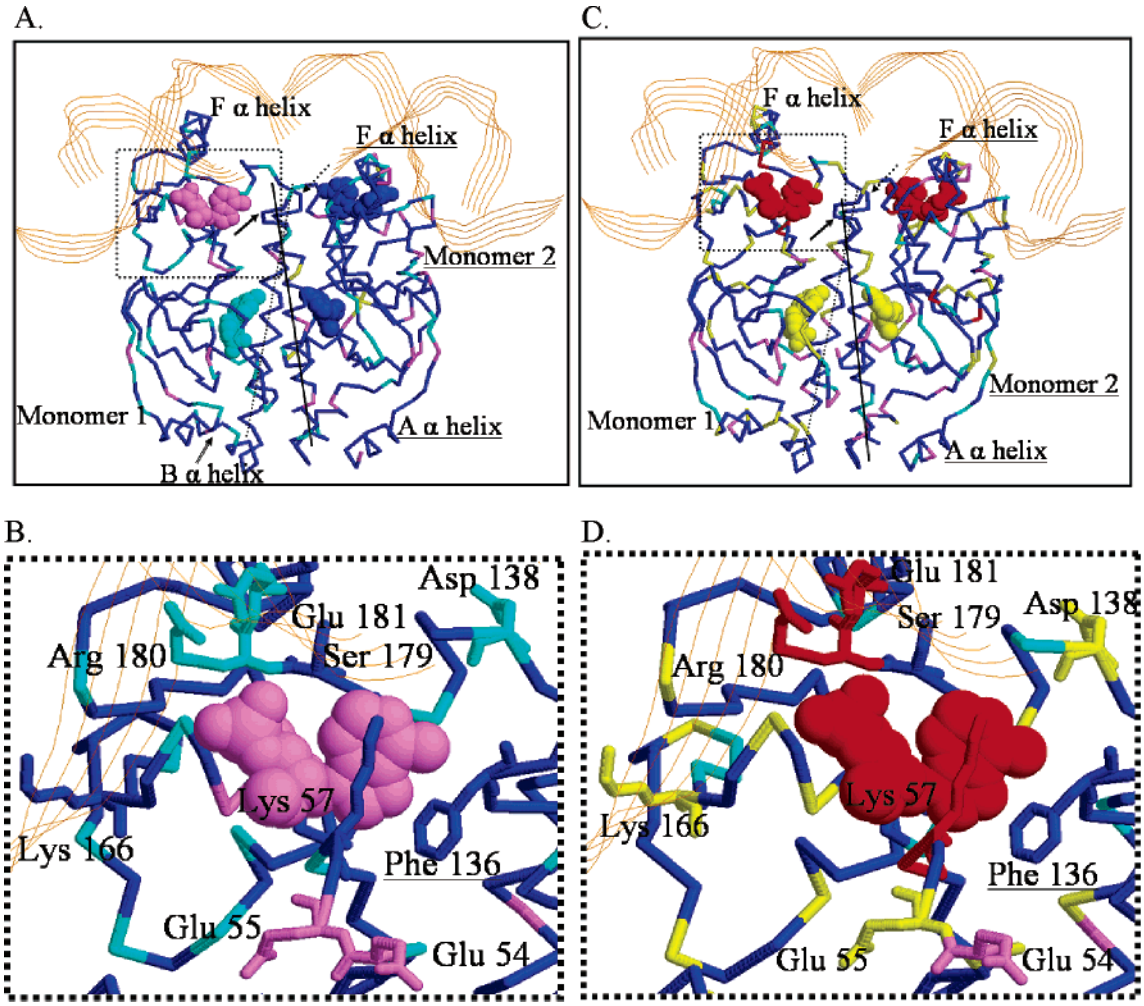

FIGURE 4: Molecular models suggest amino acid interaction energies are similar in CRP and R180K and that cAMP binding should be inhibited in the ADH secondary binding domain. (A) Dimer 1g6n-, 1o3q-, 1run-, 1lb2-, and 1i5z-derived CRP and mutant (ADH and R180K) with cAMP molecular models had the nonbonded energy for each amino acid determined by AMMP. The average absolute ($CRP_{average}$ − $R180K_{average}$)/2 external nonbonding energy difference for the five models is shown on the 2cgp crystal structure α-carbon trace. The labels and color code are as in Figure 1. (C) The average absolute ($CRP_{average}$ − $ADH_{average}$)/2 external nonbonding energy difference for the five models is shown on the 1o3q-derived CRP molecular model α-carbon trace. Panels B and D are close-ups of the dotted box regions of panels A and C, respectively.

Final molecular models for ADH, R180K, and CRP were derived from crystal structures 1g6n (*23*), 1run (*24*), 1lb2 (*25*), 1i5z, and 1o3q (*8*) from the Protein Data Bank. A similar DNA template was present during the structure determination of DNA containing crystal structures and provides the main reason for choosing these structures. The external nonbonded energies were determined for each amino acid in the models and averaged together for the five ADH-, R180K-, or CRP-derived models. There are only five amino acids with external nonbonded energy differences greater than 10 kcal/mol when the R180K- and CRP-derived models are compared (Figure 4A), and all are in the second monomer. Residue 334 (Glu 34) is located at the beginning of β-strand 3, residue 378 (Glu 78) at the beginning of β-strand 7, residue 429 (Glu 129) in the α-helix C, residue 452 (Lys 152) just after α-helix D, and the expected residue 480 (Arg/Lys 180) at the start of α-helix F. Residues Glu 129 and Lys 152 are across from each other and may position the DNA binding domain. Arg 180 interacts with the bound ligand's ribofuranose phosphate, and the difference can be attributed in part to the mutation from Arg to Lys, although the energy difference occurs in only one of the monomers.

The external nonbonded energy differences of >10 kcal/mol are dominated by a single model with a difference of more than 40 or two models with differences of greater than 20 which pushes the average differences just above the cutoff, the largest average difference being 12. There are still three differences if the threshold is dropped to 5 kcal/mol: residues 378 (Glu 78), 452 (Lys 152), and 480 (Arg/Lys 180). The threshold would have to be dropped to 2 kcal/mol for Lys 180 to show no differences between monomers (see Figure 4B), but the real difference between monomers is half the 8 kcal/mol difference between the 2 and 10 kcal/mol cutoffs. The only ligand in the R180K models that shows a difference greater than 5 kcal/mol but not greater than the 10 kcal/mol threshold is in the secondary binding domain of monomer A.

Sixty-two amino acids have external nonbonded energy differences greater than 10 kcal/mol when the average external nonbonded energies of ADH and CRP are compared (see Figure 4C). Twenty-two amino acids have differences in both monomers in the dimer, and 18 are unique to one of the monomers. The largest energy differences (>20 kcal/mol) correspond to the mutated Arg 180 and Glu 181.

There are energy differences spread throughout the models with concentrations surrounding the secondary binding sites, especially in the loop 4−5 area, the top of α-helix C, and α-helix F where the ADH mutations occur. Amino acids Lys 44 between β-strands 3 and 4; Lys 52, Glu 55, Lys 57, and Glu 58 on β-strands 4 and 5 and the intervening loop; Glu 81 and Arg 82 on β-strand 7; Arg 87 and Lys 89 on β-strand 7; Lys 101 and Lys 130 on α-helix C; Asp 138 between helices C and D; Arg 142 in α-helix D; Lys 152 and Asp 155 between α-helix D and β-strand 9; Lys 166 in β-strand 9; Arg 169 in α-helix E; Arg/Asp 180, Glu/His 181, Lys 188, and Lys 201 in β-strand 12; and Arg 209 after β-strand 12 show differences in both of the monomers making up the dimer. Residues Asp 8, Ser 16, Asp 53, Lys 100, Arg 103, Asp 111, Arg 123, Asp 161, Glu 171, Glu 191, and Asp 192 have large differences only in monomer A, while residues 322 (Lys 22), 326 (Lys 26), 335 (Lys 35), 354 (Glu 54), 393 (Glu 93), 429 (Lys 129), and 485 (Arg 185) have differences in monomer B.

Clearly, the mutations have distant effects throughout the CRP dimer, and it is suggestive that the main effect of the ADH mutations is a disruption of the charged amino acid network encompassing the dimer complex. This might be exaggerated with water not included in the models, although one of the main assumptions is that water and entropic effects would be negligible when averaging the five crystal structures and subtracting the model systems from each other. Supporting the idea of a charge network is the comparative lack of differences between CRP and R180K where the charged species is conserved and the energy differences present in the CRP and R180K comparison are also present in comparison of R180K and ADH. There are 24 amino acids with different predicted energy differences in the ADH versus R180K (not shown) and ADH versus CRP comparisons. This decreases to differences in two residues, at positions 308 (Asp 8) and 337 (Glu 37), if the threshold is dropped to 5 kcal/mol. Further, all the residues which differ in external nonbonded energies when cAMP- and cGMP-bound models are compared are included in the CRP and mutant model comparisons.

The large ADH and CRP differences which face the binding pocket, Lys 57, residue 180, and residue 181, occur in both monomers (Figure 4D), leading to large differences between the bound cAMP ligand. The molecular modeling agrees with the tritium binding and proteolysis results of the ADH mutations inhibiting the secondary binding domain and is consistent with cGMP (Figure 1B) not being able to bind the secondary binding domain. The primary binding site ligands show CRP and ADH differences, but this is at the low end of the 10−19 kcal/mol range with just 11 and 13 kcal/mol.

The lack of proteolytic protection in ADH is related to the position of Lys 57. Lys 57 is crucial for determining the level of exposure of Phe 136 to solvent and therefore to chymotrypsin cleavage as suggested by the molecular models. The CRP-derived 1run model has Lys 57 across the purine face and interacting with DNA protecting Phe 136 from cleavage (Figure 5A). α-Helix C is positioned beneath the Phe ring (in black), and Asp 53 is centered over the Phe-conjugated ring system. Lys 57 interacts directly with Asp 53 above Phe 136 with bound cAMP restricting the access of Phe 136 to solvent in the R180K-derived 1run model (Figure 5B). The Phe 136 conjugated ring system is oriented toward the cAMP purine ring in response to the repositioning of Lys 57 and is no longer oriented toward Gly 132 and α-helix C. Phe 136 is oriented toward solvent (between labels Lys 57 and Phe 136) in the absence of cAMP with Asp 53 and Lys 57 still interacting, but no longer over the Phe ring in the ADH-derived 1g6n model (Figure 5C). The orientation of the Phe 136 ring is nearly parallel to that of Gly 132. The 1g6n models represent the best opportunity for chymotrypsin to bind Phe 136 and cleave the hinge region. Bound DNA protects Phe 136 from cleavage by also restricting access to the solvent (Figure 5D). Lys 57 can interact more directly with DNA in the absence of cAMP, and Asp 53 is positioned above the Phe ring. The orientation of Phe 136 with respect to α-helix C nearly matches that seen with bound DNA and cAMP.

DNA binding experiments were performed to determine the effect of the ADH and R180K mutations on their ability to selectively and nonselectively bind DNA. CRP unexpectedly showed the ability to bind the *dsrA* promoter selectively in initial DNA binding experiments using six DNA fragments (*mcl* promoter, *dsrA* promoter, *lac* promoter, repressor, LeuO, and T7 polymerase) (Figure 6A). The *dsrA* promoter from *E. coli* disappeared before the concentration of 200 nM CRP was reached in 200 μM cAMP. The *lac* promoter from the vector pAlter-1 fragment had been selectively bound by 1000 nM CRP. Nonselective DNA binding started occurring at 500 nM CRP with the largest T7 polymerase DNA and was nearly complete by 10 μM CRP monomer with the disappearance of the *mcl* promoter. This result is significant for two reasons: (1) the unexpected result of the selective binding of the *dsrA* promoter by CRP and (2) the fact that CRP has a higher apparent affinity for the selective binding of the *dsrA* promoter than for the *lac* promoter.

The *lac* promoter was replaced in the DNA fragment mix with a DNA fragment from the CRP clone that is nearly the same size. CRP selectively bound the *dsrA* promoter with a visible shift in size with the CRP-bound *dsrA* promoter in the presence of 2 mM cAMP (Figure 6B). Half the total *dsrA* promoter DNA was bound between 20 and 200 nM CRP and completely bound by 300 nM CRP. The nonselective DNA binding starts at 300 nM CRP with the disappearance of the T7 polymerase DNA and is completed by 5 μM with the disappearance of the *mcl* promoter DNA. R180K with 2 mM cAMP visibly selectively binds the *dsrA* promoter at a monomer concentration of 2 μM, while nonselective DNA binding starts occurring at a monomer concentration of 1 μM (Figure 6C). The R180K nonselective DNA binding is complete with the disappearance of the *mcl* promoter by 5 μM. ADH with 2 mM cAMP shows no visible selective DNA binding and minimal nonselective DNA binding starting at 2 μM monomer (Figure 6D). Only the large T7 polymerase DNA fragment is absent once the monomer concentration reaches 9 μM with minimal disappearance of the LeuO DNA fragment. The same batch of DNA fragment mixture was used in the experiments shown in Figures 6B−D. Experiments performed with 200 μM and 20 mM cAMP showed no visible change in nonselective DNA binding and changes that were hard to quantitate for selective DNA binding. Further, the results with the *lac* promoter mirror those with the *dsrA* promoter except that R180K shows no selective *lac* promoter DNA binding. A different

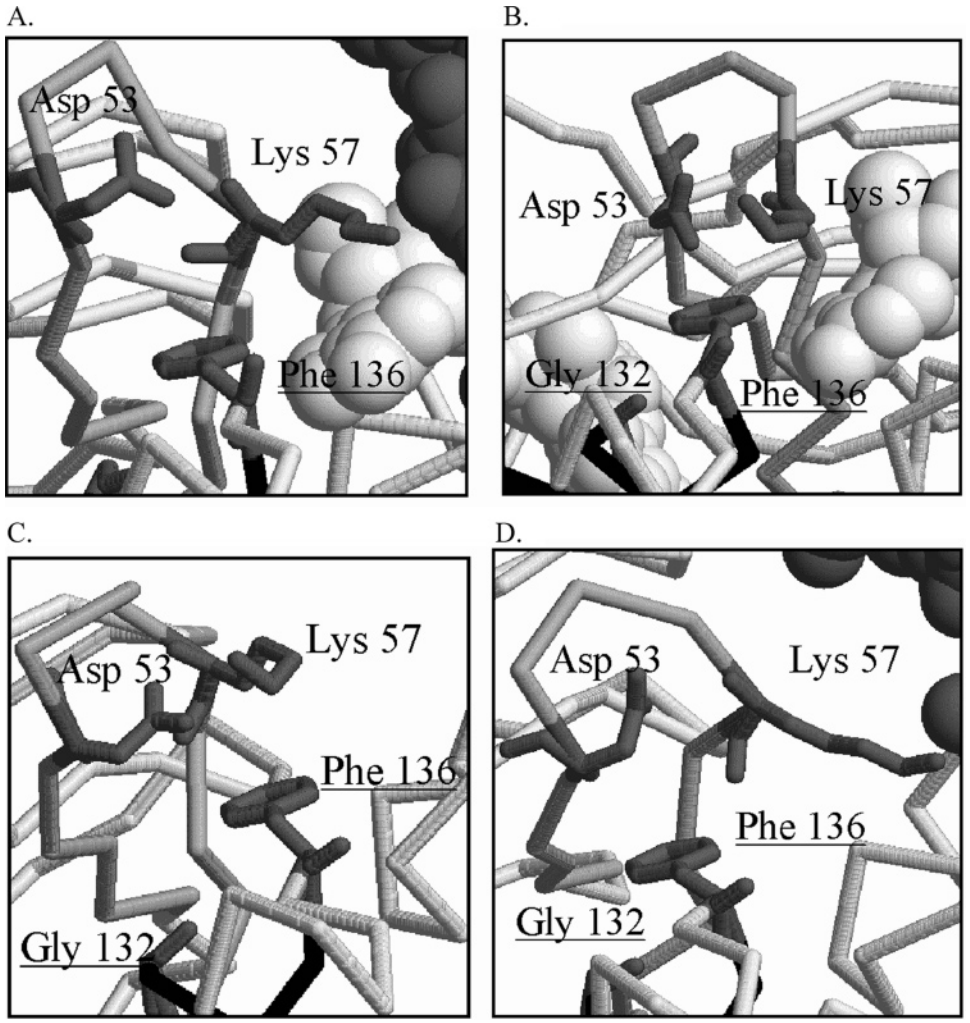

FIGURE 5: Molecular models imply that limited solvent exposure of Phe 136 inhibits chymotrypsin cleavage. (A) Asp 53, Lys 57, and Phe 136 1o3q-derived CRP molecular model interactions with bound DNA (gray space filling) and cAMP (white space filling). The $\beta 4-\beta 5$ loop α-carbons (light gray) and α-helix C (black) are highlighted. The color scheme is used in the rest of Figure 5. (B) Asp 53, Lys 57, Gly 132, and Phe 136 1i5z-derived R180K molecular model interactions with cAMP in the absence of DNA. (C) Asp 53, Lys 57, Gly 132, and Phe 136 1g6n-derived ADH molecular model interactions in the absence of DNA and cAMP. The area between the Lys 57 and Phe 136 labels is solvent-exposed. (D) Asp 53, Lys 57, Gly 132, and Phe 136 1run-derived CRP molecular model interactions with bound DNA in the absence of cAMP.

experimental method needs to be employed to determine the ligand-dependent changes in selective DNA binding for R180K and the *dsrA* promoter and exhaustive comparisons of the relative affinities of CRP for *lac* and *dsrA* promoters.

The inability of R180K to selectively bind the *dsrA* promoter compared to CRP is difficult to understand without the use of molecular models described earlier. The 1run crystal structure (Figure 7B) shows that Arg 169 (on α-helix F in black) and Arg 180 and Glu 181 (on α-helix E in dark gray) interact with DNA (black ribbon). Arg 180 and Glu 181 face each other with an axis nearly parallel to α-helix E. Arg 169 and Arg 180 are drawn to the bound cAMP phosphate, effectively changing the axis between Arg 180 and Glu 181 to a near-perpendicular position in the 1o3q crystal structure (Figure 7B). CRP 1o3q- and 1run-derived molecular models show a similar trend with the initial position of the Arg 180−Glu 181 axis nearly perpendicular moving toward a horizontal position 180° from the 1run crystal structure (Figure 7C,D). Arg 169 is also drawn to cAMP's phosphate in a motion more exaggerated than that seen in the crystal structure. The R180K 1o3q- and 1run-derived models show Arg 169, Lys 180, and Glu 181 motions similar to those of the CRP-derived models (Figure 7E,F). The Lys 180 point charge in R180K interacting more directly with Glu 181 rather than stacking with Glu 181 like Arg 180 in CRP is the major difference in the R180K models. Lys 180 is not able to interact with both the DNA and Glu 181 simultaneously. The lack of crystallographic and non-crystallographic water is the cause of the exaggerated motions seen in the molecular models.

ADH's loss of nonspecific DNA binding can also be explained with the molecular models. The ADH 1o3q-derived model displays interactions between residues Arg 169, Asp 180, and His 181 and bound DNA (Figure 8A). This model differs from the CRP and R180K 1o3q-derived models in that both Asp 180 and His 181 interact with the same strand of DNA, but this would likely not explain the near-total loss of nonselective DNA binding. The ADH 1g6n-derived model provides a possible explanation (Figure 8B). Asp 180 directly interacts with Arg 169, likely precluding DNA from disrupting the interaction and therefore precluding DNA binding. The CRP 1g6n-derived model

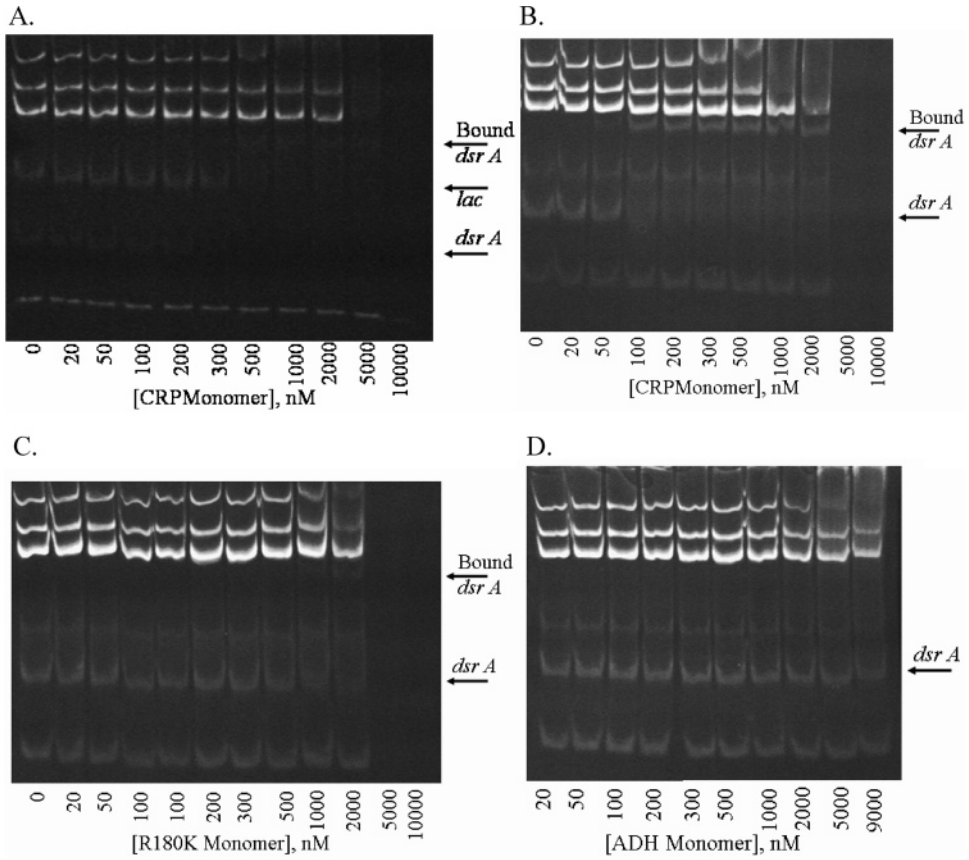

FIGURE 6: R180K selective and ADH nonselective DNA binding is disrupted. (A) The CRP *dsrA* promoter affinity is greater than the *lac* promoter affinity. The early disappearance of the *dsrA* and *lac* promoter DNA fragments denotes selective DNA binding. Nonselective DNA binding occurs when the other DNA fragments disappear. The cAMP concentration is 200 μM. (B) The apparent molecular weight of the *dsrA* promoter fragment shifts with bound CRP. The CRP/*dsrA* promoter $K_{0.5}$ is between 20 and 200 nM CRP monomer, as determined by visual inspection. Another DNA fragment replaces the *lac* promoter. The cAMP concentration is 2 mM for panels B−D. (C) R180K nonselectively binds DNA. *dsrA* promoter binding is detectable at 1000−2000 nM. (D) ADH nonselective DNA binding is detectable at 5 μM ADH monomer.

suggests a similar interaction may occur between Arg 169 and Glu 181 (Figure 8C), but this interaction is not maintained in the 1g6n crystal structure (Figure 8D). The Arg 169−Asp 180 interaction is much more likely to occur than the Arg 169−Glu 181 interaction because of the difference in position on α-helix E with Asp 180 facing Arg 169 and Glu 181 being solvent-exposed in one position further down on the α-helix. The ADH nonselective binding of DNA is likely inhibited by the Arg 169−Asp 180 interaction as DNA is not able to effectively compete with Asp 180 for a crucial interaction with Arg 169.

DISCUSSION

The CRP secondary binding sites and the residues directly impacted by bound cAMP have been explored. Binding of cAMP to the secondary binding sites inhibits selective DNA binding by repositioning residues Arg 180 and Glu 181 away from the major groove while having little effect of nonselective DNA binding by allowing Arg 169 to continue interacting with bound DNA. Secondary binding site-bound cAMP, Leu 57, and Asp 53 decrease Phe 136 solvent and protease access, protecting the hinge region from proteolytic cleavage. Finally, the secondary binding sites are capable of binding cAMP only after the primary binding sites have bound cAMP. These results are consistent with previous reports but differ by focusing on the contribution of the secondary binding sites to proteolytic protection as well as selective and nonselective DNA binding and providing further evidence for a 2 × 2-binding site ligand binding model.

The following discussion is predicated on the assumption that the functional allosteric states which CRP and the two mutants can occupy are equivalent to each other. The data support this assumption as it relates to the binding of the ligand in the primary binding site. Both mutants are capable of achieving the functional CRP states as viewed by the proteolytic cleavage of Phe 136. The mutants differ from CRP in that R180K has inhibited selective DNA binding due to the point charge of lysine versus the dispersed charge of arginine. ADH differs from R180K and CRP in that the mutations are incapable of binding the cAMP ribofuranose, resulting in one ligand binding per monomer. The mutations further remove the necessary residues for interaction with the DNA major groove and therefore remove DNA selective binding. An additional effect is that nonselective DNA binding is compromised due to intramolecular interactions between the mutated R180D and Arg 169. Even with these differences, the nearly equivalent ligand binding and proteolytic cleavage suggest the three proteins go through the same allosteric states even though the mutations preclude function from being carried out.

The 2 × 2-binding site model is dependent on the primary binding domains being occupied before the secondary

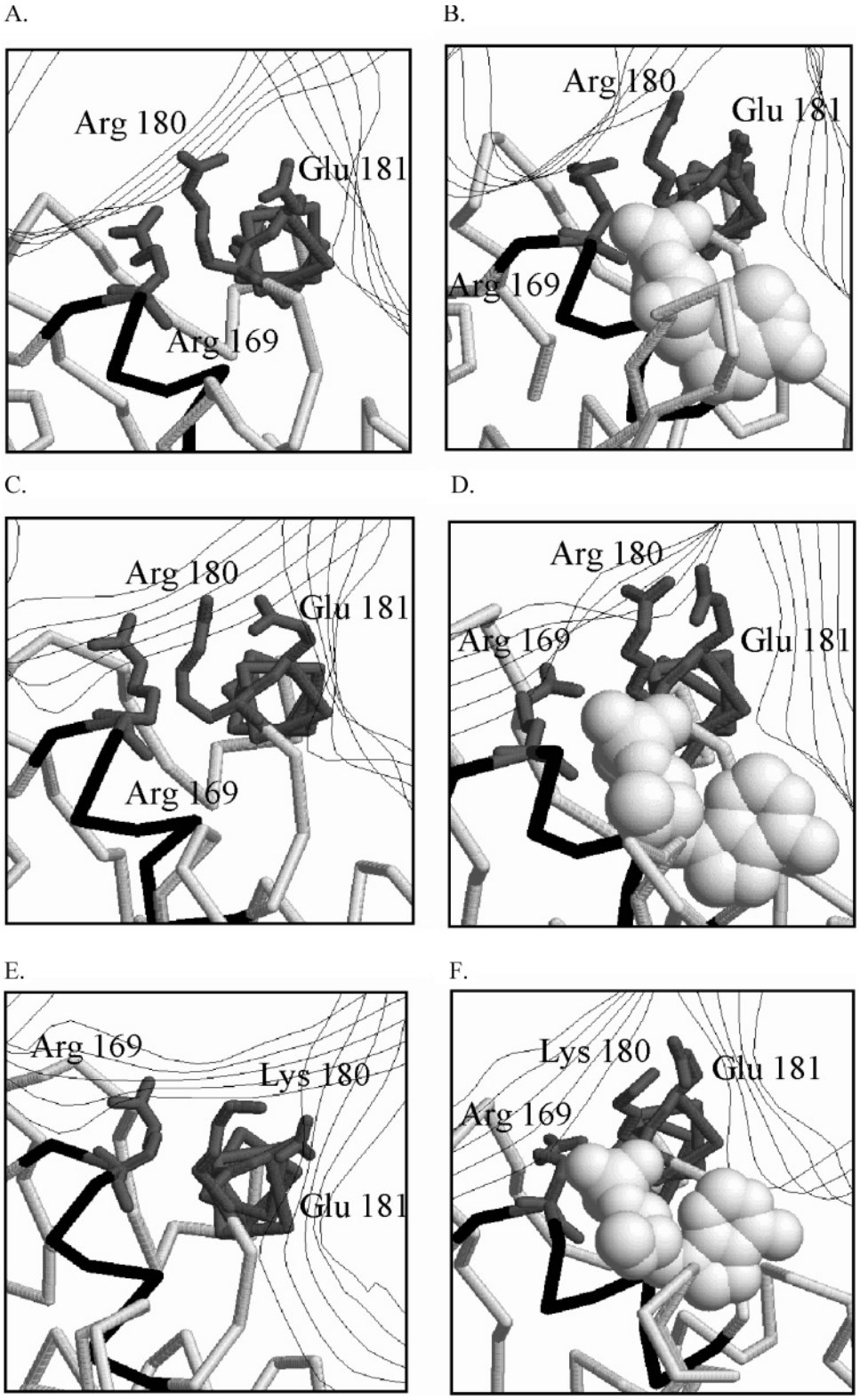

FIGURE 7: Molecular modeling insinuates selective DNA binding is dependent on Arg 180 positioning Glu 181. (A) Arg 169, Arg 180, and Glu 181 (gray) 1run crystal structure interactions in the presence of DNA (black strands). α-Helix E (black) and α-helix F (gray) are highlighted. The color scheme is used in Figures 7 and 8. (B) Arg 169, Asp 180, and His 181 1o3q crystal structure interactions in the presence of DNA and cAMP. (C) Arg 169, Arg 180, and Glu181 1run-derived CRP molecular model interactions in the presence of DNA. (D) Arg 169, Arg 180, and Glu 181 1o3q-derived CRP molecular model interactions in the presence of DNA and cAMP. (E) Arg 169, Lys 180, and Glu181 1run-derived R180K molecular model interactions in the presence of DNA. (F) Arg 169, Lys 180, and Glu 181 1o3q-derived R180K molecular model interactions in the presence of DNA and cAMP.

binding domains in a sequential pattern. There are two chances for this kinetic mechanism to occur: (1) the affinities of the primary binding domains, on the order of 10 000 M^{-1}, are much greater than those of the secondary binding domains, on the order of 1000 M^{-1}, such that mass action dictates they are bound first, or (2) the primary binding

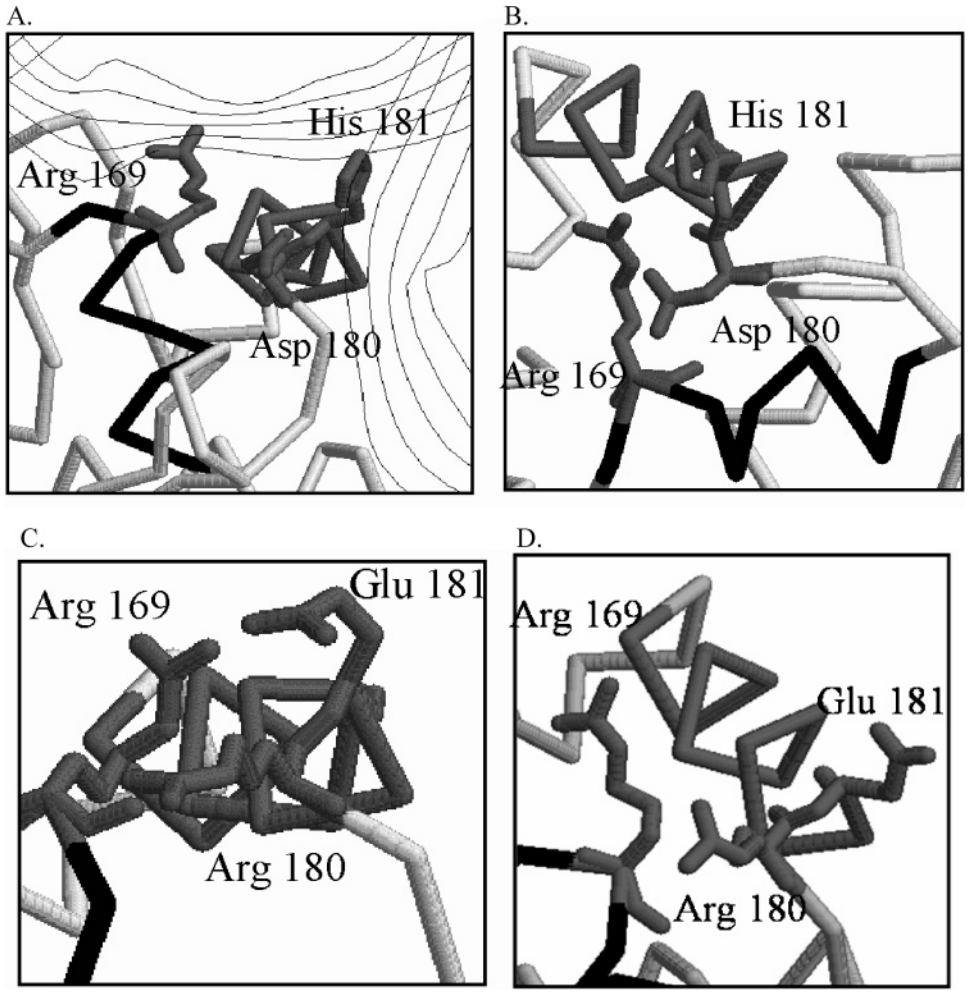

FIGURE 8: Molecular modeling suggests Arg 169 contributes to nonselective DNA binding. (A) Arg 169, Asp 180, and His 181 1run-derived ADH molecular model interactions with bound DNA. (B) Arg 169, Asp 180, and His 181 1g6n-derived ADH molecular model interactions in the absence of DNA. (C) Arg 169, Arg 180, and Glu181 1g6n-derived CRP molecular model interactions in the absence of DNA. (D) Arg 169, Arg 180, and Glu 181 1g6n crystal structure interactions in the absence of DNA.

domains must be filled first for the secondary binding domains to be formed and therefore filled. The first possibility is extremely difficult to refute because all data are supportive of the primary binding domains having a much higher affinity for ligand than the secondary domains in vitro. The situation in vivo is much the same, with an initial activation at low concentrations of cAMP followed by an inhibition at cAMP concentrations more than 10-fold higher. Further, the in vivo CRP concentration is estimated to be 2.5 μM (*27*). suggesting that a significant amount of the protein will be complexed with DNA in the absence of cAMP. The in vivo CRP−DNA complex has two consequences: (1) physiological concentrations of cAMP (<1 μM) are able to control CRP promoter binding and therefore activation (*12*), and (2) the secondary sites would already be formed and capable of binding cAMP as represented by the six crystal structures used for the molecular modeling experiments. These consequences would support a kinetic mechanism in which the higher affinity of the primary sites leads to binding of cAMP to them prior to secondary sites.

The CRP conformation in the absence of DNA or cAMP is unknown. The primary evidence that the secondary binding domains are unformed in the absence of DNA or cAMP is the crystal structure of another member of the CRP−FNR transcription factor family, CooA from *Rhodospirillum rubrum*. CooA contains a heme fixed in what would be CRP's primary binding domain and responds in vivo to CO concentration changes. The CooA crystal structure has one of the DNA binding domains in a near-parallel plane with the primary binding domain (*40*) as opposed to the perpendicular plane seen in all CRP crystal structures. The second DNA binding domain is positioned between the parallel binding domain and the most parallel CRP DNA binding domain represented by the "open" conformation of the 1g6n−CRP crystal structure (*6*, *23*).

Independent sources of evidence suggest that CRP is capable of forming a structural conformation similar to CooA with the DNA binding domains parallel to the primary binding domains. This parallel conformation implies that the secondary binding domains are unformed in the absence of ligand or DNA and would need to be formed. The first evidence of a large structural conformational change is in microcalorimetry experiments (*11*), where cAMP binding to the secondary site includes an entropic penalty. This implies that the DNA binding domain is flexible in solution and becomes more ordered upon binding of cAMP.

This entropic penalty might prove so great that the CRP dimer is incapable of binding a fourth cAMP molecule in the absence of DNA as supported by the 1i5z crystal struc-

ture. This is in contrast to the tritium binding data which suggest a strong positive cooperativity between the two secondary sites but more consistent with the assertion that the secondary sites experience negative cooperativity (*12*). This could suggest a 2 × 1-site model is more appropriate in the absence of DNA and that the cooperativity shown in the CRP and R180K tritium data is an anomaly of forcing a fourth molecule to bind. The data were not of sufficiently high quality to distinguish between the 2 × 1- and 2 × 2-site binding models on the basis of the goodness of fit, although the fits consistently underestimated the number of bound ligands at high concentrations. The monomeric proteolytic cleavage model does not resolve this issue.

Another source of evidence comes from studies with the physiologically inactive ligand cGMP. The molecular models suggest that the purine ring is the cause of cGMP not being able to bind to the secondary binding domain. If this was the case, then chymotrypsin should completely cleave the hinge region separating the primary binding domain from the DNA binding domain. Instead, chymotrypsin is unable to cleave the hinge region which is much like what occurs in the absence of any ligand, implying cGMP is incapable of forming the perpendicular DNA binding domain structural state necessary to expose the hinge region to solvent and chymotrypsin cleavage. The inability of cGMP to form the perpendicular structural state would also explain why cGMP is incapable of binding more than two molecules per CRP dimer, promote selective DNA binding, or activate transcription.

Many of the CRP mutants [D53H, S62F, G141D, R142C, and L148R (*41*), T127C, T127S, and T127I (*42*), G141S, G141R, G141K, G141D, A144S, A144Q, A144Y, A144L, A144F, A144V, and A144C (*43*), G141Q (*43*−*45*), and A144T (*46*−*48*)] capable of exhibiting activity with cGMP show either a sensitivity to proteolytic cleavage or *lac* promoter activity in the absence of cAMP (*23*), suggesting that the cAMP allosteric motions necessary for transcription or proteolytic cleavage have already partially or fully formed without regard to bound ligand. The three mutants [T127L (*49*), S128T (*42*), and D138N (*50*)] in which cGMP is able to activate transcription and which are not sensitive to protease cleavage or activity in the absence of cAMP might remove the selectivity for cAMP and just need the presence of the ribofuranose to position the $\beta4-\beta5$ loop for activity while α-helix C is either already correctly positioned or able to be positioned by cGMP. T127L further shows the inability to be activated by bound cAMP, suggesting that it is incapable of correctly orienting α-helix C. Testing the cGMP active mutant's secondary binding site's ability to bind cGMP would determine whether they conform to a 2 × 2-binding site model and whether cGMP is capable of binding to the CRP secondary binding sites.

The CRP with cGMP results correspond with the cAMP-bound ADH results in binding two molecules per dimer, not selectively binding DNA, and being unable to promote transcription. They differ on the important point of chymotrypsin cleavage. ADH mimics the initial stages of CRP cAMP-dependent protein cleavage by promoting cleavage of the hinge region, implying that ADH is capable of structural motions leading toward Phe 136 exposure. This structural conformation would allow ADH to selectively bind DNA and promote transcription if not for the mutations in the DNA binding domain as ADH nonselectively binds DNA with significantly lower affinity than CRP without any ligand (*30*) or presumably with cGMP. This motion is cAMP-dependent and different from the cGMP structural states as cGMP does not promote ADH hinge cleavage. Molecular modeling comparisons with cAMP-bound CRP suggest cGMP is more comfortable in the CRP primary binding site than cAMP is in the ADH primary binding sites. This suggests not only the interdependence of all residues in CRP but also that the structural state of bound cGMP must be fundamentally different compared to that of cAMP or the hinge region would be cleaved. The most likely structural state is one that mimics the CooA crystal structure with α-helix C, the hinge region, and α-helix D forming one long α-helix held together by interactions with the dimer's second α-helix. cAMP in CRP and ADH is able to promote α-helix breakage, whereas cGMP is not.

The transition from the parallel to the perpendicular structural state is likely caused by interactions of cAMP with S128 on α-helix C of the second monomer. This interaction likely breaks helix−helix interactions by rotating α-helix C compared to the eight-stranded β-barrel of the primary binding domain. The second α-helix C would respond in kind, explaining the positive cooperativity seen between the two primary binding domains as a second cAMP ligand would stabilize the state. The result would be the DNA domain swinging into the perpendicular position by the unfolding of large α-helix C−D into the resulting α-helix C, hinge region, and α-helix D. This realignment of the DNA binding domain would expose Phe 136 to chymotrypsin cleavage, allow CRP to bind DNA and promote transcription, and form the secondary binding domains.

The syn cAMP secondary binding site was previously defined as residues 56−59, 170, 174, and 177−180 (*9*). The slight decrease in the level of nonselective DNA binding in the presence of high cAMP concentrations suggests that Arg 169 should be included in this list because of its predicted interaction with cAMP's ribofuranose. The R180K mutant's inability to selectively bind DNA suggests that the orientation of Glu 181 is dependent on the orientation of Arg 180 which is further dependent on the presence of cAMP in the secondary binding site. Therefore, Glu 181 should be considered part of the extended binding domain defined as those amino acids which do not directly interact with bound ligand but are directly affected by the presence of ligands. The proposed interaction between Arg 180 and Glu 181 is consistent with the crystallography observations that both residues are in the major groove (*10*) and that amino acid 181 provides an essential interaction when determining DNA specificity (*24*). The partial positive charge of the Phe 136 ring interacts with the partial negative charge of the cAMP purine ring and therefore should be considered part of the secondary binding site binding domain proper, defined as the amino acids that interact directly with a bound ligand. The Asp 53 position is directly related to the position of Leu 57 and Phe 136 and should be considered part of the extended secondary binding site domain. The positioning of Asp 53 and Leu 57 on the $\beta4-\beta5$ loop and Phe 136 in the hinge region near the end of α-helix C by binding of cAMP to the primary binding sites forms the cAMP secondary binding site.

The secondary binding site being composed of residues from both monomers is likely the "sensor" for showing that

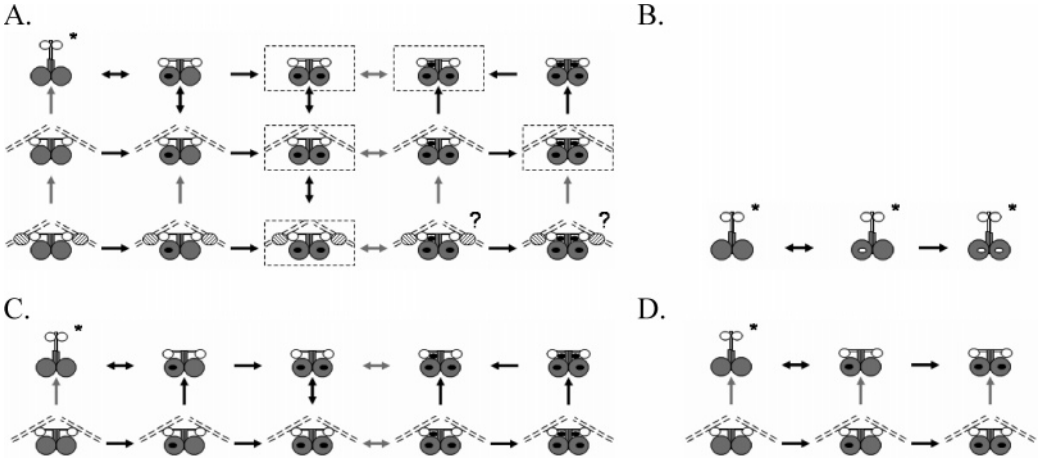

FIGURE 9: R180K can occupy 10, ADH six, and cGMP three of the 15 speculative CRP allosteric protein states. (A) CRP protein (gray oval representing the cyclic nucleotide binding domain, a gray and square box representing α-helices C and D, and the DNA binding domain represented by a white oval) state diagram with zero to four bound cAMP ligands (black ovals), zero to one bound DNA fragment (dotted parallel lines), and zero to one bound RNA polymerase (oval with diagonal lines, the presence of which denotes selective DNA binding). Dark lines mark paths between energetically favorable states, and the arrowhead points to the favored state. Energetically costly paths are shown as gray lines. The asterisk state is a hypothetical state in the absence of ligand. Two states which may be possible are marked with ?. The dotted boxes are the five crystallized CRP states. (B) cGMP (white ovals) in CRP, ADH, and R180K would be able to reach the same three hypothetical allosteric states but is unable to break the C−D α-helix at the hinge region when bound. (C) R180K can achieve the 10 protein states. (D) ADH is confined to six protein states.

both primary binding sites are bound and allowing the secondary binding sites to now be occupied. This is consistent with the crystallographic data with both primary binding sites being occupied before either secondary binding site is occupied (*9*, *30*) and the similar binding constants and binding curves of ADH and the occupation of the primary sites of CRP and R180K. What is unclear is whether both secondary binding sites will bind ligand in the absence of DNA and whether there is cooperativity between the binding sites or whether all the measurements are just assessing the binding of ligand to the first of the secondary binding sites. The agreement of the proteolytic assay association constants with the CRP tritium binding association constant suggests that a single ligand in either secondary binding site may be enough to protect Phe 136 from chymotrypsin cleavage. The large errors at high concentrations of the tritiated ligand mean the experiments presented here are not useful in clarifying the issue.

It is clear that four ligands per CRP dimer is likely a common state given that the DNA "fragments" present in vivo would be larger than the 30 bp fragments present in crystal structures which lack cAMP in the secondary sites. Therefore, CRP would be capable of binding four ligands per dimer and not limited to binding two ligands. What is not clear is whether a third ligand per dimer is necessary for selective DNA binding, although it is clear that a fourth ligand would decrease the level of selective DNA binding in the *lac* promoter given the inhibitory effect of cAMP at high ligand concentrations. The chymotrypsin data suggest that a third ligand inhibits access to the hinge region, but similar results occur with DNA present. The proteolytic data could be measuring the extent of binding of the ligand to the fourth site, as could the tritium data given the poor quality of data at high [^{3}H]cAMP concentrations. Only the difficulty in obtaining crystallographic data with three bound ligands, the exceedingly high concentrations necessary for proteolytic cleavage protection (the half-inhibition concentration between 2 and 5 mM), and the low physiological cAMP concentration suggests that the third binding site is only commonly occupied in the presence of DNA. The necessity of DNA for cAMP secondary site occupation is indicative of a purely inhibitory role in selective DNA binding.

Finally, DNA binding experiments were originally conducted with *lac* as the positive control and *dsrA* as a negative control. These experiments showed that the CRP selectively bound the *dsrA* promoter at protein concentrations 10-fold lower than that needed to bind the *lac* promoter. Further, the *dsrA* binding seemed to be robust even at high concentrations (20 mM) of ligand, although these were not the proper experiments for truly testing the ligand concentration dependence of binding of CRP to the *dsrA* promoter. The result implies that CRP is capable of binding to the *dsrA* promoter and may be the hypothesized unknown transcription factor. The DNA region from position −75 through −53 (ATTCATAAGTA·GCGTTAATCAT) of the *dsrA* promoter is a likely candidate for the CRP binding region when compared to a consensus CRP binding sequence (AAATGTGATCT·AGATCACATTT) and a *lac* DNA binding site (TAATGTGAGTT·AGCTCACTCAT) (*51*). This corresponds with the loss of a weak activation element in expression by the *dsrA* promoter when the DNA region between base pairs −165 and −64 is removed from the promoter (*19*). Further experiments need to be performed to definitively determine if CRP is truly the control protein which binds to the *dsrA* promoter and if this is the CRP binding site.

A speculative allosteric model emerges for the different CRP protein states composed of 15 protein states (Figure 9A). Each protein state is different from the others in the number of ligands bound (top row), whether DNA is nonselectively bound (middle row) or DNA is selectively bound (bottom row). Five of the protein states have been crystallized (dotted box) and represent energetic minima. A "?" marks two protein states as no crystallographic or functional data support the ability of CRP to effectively selectively bind DNA and promote transcription in the presence of four cAMP molecules.

The asterisked protein state could have a structure more similar to CooA structure with the DNA domain parallel to the primary ligand binding domain and not perpendicular (*40*). The black lines represent lower-energy paths that are more likely to occur unless large amounts of ligand and DNA are provided in relation to the concentration of CRP. The arrowhead shows where the equilibrium state is most likely to lie. This allosteric model is an attempt to organize the major CRP protein states and is not an attempt to determine the association constants or the mechanistic steps leading to each state. The pathways and arrows are the suggested paths, and relative strengths and directions and are not intended to represent equilibrium constants or real minima. The CRP structures which have been crystallized are assumed to be energy minima under the conditions of crystal growth.

CRP is able to occupy all 15 states, preferring the DNA-free states until a ligand is bound, and then a DNA-bound state is preferred in vitro. The cooperativity of binding sites 1 and 2 suggests that once a single ligand is bound a second ligand binds quickly to stabilize the CRP structure as seen in the middle column of Figure 9A. The second bound ligand also stabilizes the protein state necessary for selective DNA binding. The transition to three bound ligands is energetically expensive due to the low affinity of the secondary binding sites for cAMP, although the affinity could increase in the presence of DNA. CRP prefers a protein state that contains nonselective DNA once a third ligand is bound, preferring that the second secondary binding site is filled. CRP prefers to bind no more than two ligands when DNA is absent. CRP in vivo would prefer to occupy the second row of allosteric states due to the high relative DNA and protein concentrations. An increasing cAMP concentration would stabilize movement to the third row, denoting selective DNA binding and the ability to promote transcription. The energetic penalty for binding the third cAMP molecule could be reduced as the large entropic penalty would not apply because the DNA binding domain is already ordered.

R180K is limited to the first two rows that do not include selective DNA binding and therefore is unable to promote RNA transcription. R180K can freely nonselectively bind DNA and bind any number of ligands with nearly the same affinity as CRP (Figure 9C). The inability to selectively bind DNA keeps it from replacing CRP as an effective transcription factor but does not limit it from binding DNA and possibly interrupting other bacterial functions. ADH is limited to six protein states and, in reality, only the three protein states in the first row, as the nonselective DNA binding is inhibited by the mutations it carries (Figure 9D). ADH is incapable of effectively binding DNA and can therefore cause no interference with other cellular functions except those dependent on cyclic nucleotide changes as the molecule is an effective sink of both cAMP and cGMP, having affinities near those of R180K and CRP.

cGMP with CRP would be limited to the first row of the CRP allosteric model (Figure 9B). cGMP is not able to form the allosteric state necessary for nonselective or selective DNA binding. cGMP is also incapable of forming the secondary binding site when the primary site is occupied. This is more limiting than ADH as ADH is capable of forming the secondary binding sites and allosteric state necessary for DNA binding but the mutated amino acids preclude both. The inability of cGMP to form the secondary binding site and the DNA binding allosteric state is indicative of cAMP interactions with the $\beta4-\beta5$ loop and α-helix C being crucial for promoting transcription.

Several physiological roles for the secondary binding sites can be posited. (1) Inhibiting selective DNA binding and in turn the amount of transcript made at high cAMP concentrations is one role. (2) Stabilizing a CRP−DNA complex is another role, reducing from three to two the dimensional space that has to be searched by CRP. The loss of DNA selectivity allows CRP to walk along the *E. coli* genome until a selective DNA sequence can effectively compete with the bound cAMP ligand. (3) Keeping CRP primed and ready to occupy a site vacated by CRP and RNA polymerase to start a new RNA transcript is the third role. (4) Cyclic AMP bound to the secondary binding sites might favor certain promoter regions over the *lac* promoter as the *lac* promoter selectivity is altered but the nonselective component is retained until much higher cAMP concentrations are reached. These secondary cAMP binding site roles depend on the effectiveness of DNA competition with the secondary site bound cAMP for the DNA binding domain. The secondary binding sites likely help localize CRP to DNA, then allow CRP to move along DNA until a promoter region is found, and occupy the promoter region by effective competition of a DNA sequence for the CRP DNA binding domain once the site is vacated by RNA polymerase and a different CRP molecule. Low cAMP concentrations would promote free CRP concentrations that exist in lower energetic states unable to selectively or nonselectively bind DNA and therefore promote RNA transcription.

ACKNOWLEDGMENT

We thank Drs. Stuart E. Dryer at the University of Houston and Jacqueline C. Tanaka at Temple University for providing the space to conduct experiments and discussions of this work. We thank Drs. Joseph Jez at Donald Danforth Plant Science Center (St. Louis, MO), Irene Weber at Georgia State University (Atlanta, GA), and Gregg Wells at the Texas A&M University System Health Science Center (College Station, TX) for comments on the manuscript and valuable discussions pertaining to this work. We thank Patrick W. Shea for proofreading the final draft of the manuscript.

REFERENCES

1. Botsford, J. L., and Harman, J. G. (1992) Cyclic AMP in Prokaryotes, *Microbiol. Rev. 56*, 100−122.
2. Busby, S., and Ebright, R. H. (1999) Transcription activation by catabolite activator protein (CAP), *J. Mol. Biol. 293*, 199−213.
3. Harman, J. G. (2001) Allosteric regulation of the cAMP receptor protein, *Biochim. Biophys. Acta 1547*, 1−17.
4. Cossart, P., Gicquel-Sanzey, B., and Adhya, S. (1982) Cloning and sequence of the crp gene of *E. coli* K 12, *Nucleic Acids Res. 10*, 1363−1378.
5. Aiba, H., Fujimoto, S., and Ozaki, N. (1982) Molecular cloning and nucleotide sequencing of the gene for *E. coli* cAMP receptor protein, *Nucleic Acids Res. 10*, 1345−1361.
6. Weber, I. T., and Steitz, T. A. (1987) Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution, *J. Mol. Biol. 198*, 311−326.
7. Passner, J. M., and Steitz, T. A. (1997) The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer, *Proc. Natl. Acad. Sci. U.S.A. 94*, 2843−2847.
8. Chen, S., Vojtechovsky, J., Parkinson, G. N., Ebright, R. H., and Berman, H. M. (2001) Indirect Readout of DNA Sequence at the

Primary-kink Site in the CAP-DNA Complex: DNA Binding Specifically Based on Energetics of DNA Kinking, *J. Mol. Biol. 314*, 63−74.
9. Chu, S. Y., Tordova, M., Gilliland, G. L., Gorshkova, I., Shi, Y., Wang, S., and Schwarz, F. P. (2001) The structure of the T127L/S128A mutant of cAMP receptor protein facilitates promoter site binding, *J. Biol. Chem. 276*, 11230−11236.
10. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) Crystal structure of a CAP-DNA complex: The DNA is bent by 90°, *Science 253*, 1001−1007.
11. Lin, S.-H., and Lee, J. C. (2002) Communications between the High-Affinity Cyclic Nucleotide Binding Sites in *E. coli* Cyclic AMP Receptor Protein: Effect of Single Site Mutations, *Biochemistry 41*, 11857−11867.
12. Lin, S.-H., and Lee, J. C. (2002) Linkage of Multiequilibria in DNA Recognition by the D53H *Escherichia coli* cAMP Receptor Protein, *Biochemistry 41*, 14935−14943.
13. Lis, J. T., and Schleif, R. (1973) Different Cyclic AMP Requirements for Induction of the Arabinose and Lactose Operons of *Escherichia coli*, *J. Mol. Biol. 79*, 149−162.
14. Cheng, X., Kovac, L., and Lee, J. C. (1995) Probing the Mechanism of CRP Activation by Site-Directed Mutagenesis: The Role of the Serine 128 in the Allosteric Pathway of cAMP Receptor Protein Activation, *Biochemistry 34*, 10816−10826.
15. Angulo, J. A., and Krakow, J. S. (1985) Effect of deoxyribopolymers and ribopolymers on the sensitivity of the cyclic-AMP receptor protein of *Escherichia coli* to proteolytic attack, *Arch. Biochem. Biophys. 236*, 11−16.
16. Emmer, M., de Crombrugghe, B., Pastan, I., and Perlman, R. (1970) Cyclic AMP receptor protein of *E. coli*: Its role in the synthesis of inducible enzymes, *Proc. Natl. Acad. Sci. U.S.A. 66*, 480−487.
17. Ebright, R. H., Le Grice, S. F. J., Miller, J. P., and Krakow, J. S. (1985) Analogs of Cyclic AMP that Elicit the Biochemically Defined Conformational Change in Catabolite Gene Activator Protein (CAP) but do not Stimulate Binding to DNA, *J. Mol. Biol. 182*, 91−107.
18. Won, H.-S., Lee, T.-W., Park, S.-H., and Lee, B. J. (2002) Stoichiometry and structural effect of the cyclic nucleotide binding to cyclic AMP receptor protein, *J. Biol. Chem. 277*, 11450−11455.
19. Repoila, F., and Gottesman, S. (2001) Signal Transduction Cascade for Regulation of RpoS: Temperature Regulation of DsrA, *J. Bacteriol. 183*, 4012−4023.
20. Harrison, R. W. (1993) Stiffness and Energy Conservation in the Molecular Dynamics: An Improved Integrator, *J. Comput. Chem. 14*, 1112−1122.
21. Scott, S.-P., Harrison, R. W., Weber, I. T., and Tanaka, J. C. (1996) Predicted ligand interactions for 3′,5′-cyclic nucleotide-gated channel binding sites: Comparison of retina and olfactory binding site models, *Protein Eng. 9*, 333−344.
22. Scott, S.-P., and Tanaka, J. C. (1998) Use of Homology Modeling to Predict Residues Involved in Ligand Recognition, *Methods Enzymol. 293*, 620−647.
23. Passner, J. M., Schulz, S. C., and Steitz, T. A. (2000) Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution, *J. Mol. Biol. 304*, 847−859.
24. Parkinson, G., Gunasekera, A., Vojtechovsky, J., Zhang, X., Kunkel, T. A., Berman, H. M., and Ebright, R. H. (1996) Aromatic hydrogen bond in sequence-specific protein DNA recognition, *Nat. Struct. Biol. 3*, 837−841.
25. Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M., and Ebright, R. H. (2002) Structural basis of transcription activation: The CAP-α CTD-DNA complex, *Science 297*, 1562−1566.
26. Jones, T. A. and Kjeldgaard, M. (1997) Electron density map interpretation, *Methods Enzymol. 277*, 173−208.
27. Zubay, G. (1980) The isolation and properties of CAP, the catabolite gene activator, *Methods Enzymol. 65*, 856−877.
28. Horton, R. M., Ho, S. N., Pullen, J. K., Hunt, H. D., Cai, Z., and Pease, L. R. (1993) Gene splicing by overlap extension, *Methods Enzymol. 217*, 271−279.
29. Belduz, A. O., Lee, E. J., and Harman, J. G. (1993) Mutagenesis of the cyclic AMP receptor protein of *Escherichia coli*: Targeting positions 72 and 82 of the cyclic nucleotide binding pocket, *Nucleic Acids Res. 21*, 1827−1835.
30. Scott, S.-P., Weber, I. T., Harrison, R. W., Carey, J., and Tanaka, J. C. (2001) A functioning chimera of the cyclic nucleotide-binding domain from the bovine retinal rod ion channel and the DNA-binding domain from catabolite gene-activating protein, *Biochemistry 40*, 7464−7473.
31. Heyduk, T., and Lee, J. C. (1989) *Escherichia coli* cAMP receptor protein: Evidence for three protein conformational states with different promotor binding affinities, *Biochemistry 28*, 6914−6924.
32. Heyduk, E., Heyduk, T., and Lee, J. C. (1992) Intersubunit communications in *Escherichia coli* Cyclic AMP Receptor Protein: Studies of the Ligand Binding Domain, *Biochemistry 31*, 3682−3688.
33. Davanloo, P., Rosenberg, A. H., Dunn, J. J., and Studier, F. W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035−2039.
34. Klauck, E., Bohringer, J., and Hengge-Aronis, R. (1997) The LysR-like regulator LeuO in *Escherichia coli* is involved in the tranlational regulation of *rpoS* by affecting the small regulatory DsrA-RNA, *Mol. Microbiol. 25*, 559−569.
35. Lieb, M. (1966) Studies of heat-inducible λ bacteriophage, I. Order of genetic sites and properties of mutant prophages, *J. Mol. Biol. 16*, 149−163.
36. Hosono, K., Kakuda, H., and Ichihara, S. (1995) Decreasing accumulation of acetate in a rich medium by *Escherichia coli* on introduction of genes on a multicopy plasmid, *Biosci., Biotechnol., Biochem. 59*, 256−261.
37. Bernard, H.-U., and Helinski, D. R. (1979) Use of the λ Phage Promoter P_L to Promote Gene Expression in Hybrid Plasmid Cloning Vehicles, *Methods Enzymol. 68*, 482−492.
38. 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.
39. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Anal. Biochem. 182*, 319−326.
40. Lanzilotta, W. N., Schuller, D. J., Thorsteinsson, M. V., Kerby, R. L., Roberts, G. P., and Poulos, T. L. (2000) Structure of the CO sensing transcription activator CooA, *Nat. Struct. Biol. 7*, 876−880.
41. Aiba, H., Nakamura, T., Mitani, H., and Mori, H. (1985) Mutations that alter the allosteric nature of cAMP receptor protein of *Escherichia coli*, *EMBO J. 4*, 3329−3332.
42. Lee, E. J., Glasgow, J., Leu, S.-F., Belduz, A. O., and Harman, J. G. (1994) Mutagenesis of the cyclic AMP receptor protein of *Escherichia coli:* Targeting positions 83, 127 and 128 of the cyclic nucleotide binding pocket, *Nucleic Acids Res. 22*, 2894−2901.
43. Kim, J., Adhya, S., and Garges, S. (1992) Allosteric changes in the cAMP receptor protein of *Escherichia coli*: Hinge reorientation, *Proc. Natl. Acad. Sci. U.S.A. 89*, 9700−9704.
44. Cheng, X., and Lee, J. C. (1994) Absolute requirement of cyclic nucleotide in the activation of the G141Q mutant cAMP receptor protein from *Escherichia coli*, *J. Biol. Chem. 269*, 30781−30784.
45. Cheng, X., and Lee, J. C. (1998) Interactive and dominant effects of residues 128 and 141 on cyclic nucleotide and DNA bindings in *Escherichia coli* cAMP receptor protein, *J. Biol. Chem. 273*, 705−712.
46. Garges, S., and Adhya, S. (1985) Sites of allosteric shift in the structure of the cAMP receptor protein, *Cell 41*, 745−751.
47. Harman, J. G., KcKenney, K., and Peterkofsky, A. (1986) Structure−function analysis of three cAMP independent forms of the cyclic AMP receptor protein, *J. Biol. Chem. 261*, 16332−16339.
48. Weber, I. T., Gilliland, G., Harman, J. G., and Peterkofsky, A. (1987) Crystal structure of a cyclic AMP-independent mutant of catabolite gene activator protein, *J. Biol. Chem. 262*, 5630−5636.
49. Gorshkova, I., Moore, J. L., McKenney, K. H., and Schwarz, F. P. (1995) Thermodynamics of cyclic nucleotide binding to the cAMP receptor protein and its T127L mutant, *J. Biol. Chem. 37*, 21679−21683.
50. Ryu, S., Kim, J., Adhya, K., and Garges, S. (2004) Pivotal role of amino acid position 138 in allosteric hinge reorientation of cAMP receptor protein, *Proc. Natl. Acad. Sci. U.S.A. 90*, 75−79.
51. 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.

BI0479609