Directed Evolution of the Escherichia coli cAMP Receptor Protein at the cAMP Pocket*§

Sanjiva M. Gunasekara‡, Matt N. Hicks§, Jin Park¶, Cory L. Brooks‡, Jose Serate¶, Cameron V. Saunders‡, Simranjeet K. Grover‡, Joy J. Goto§, Jin-Won Lee**, and Hwan Youn†1

From the Departments of §Biology, ¶Computer Science, and ‡Chemistry, California State University Fresno, Fresno, California 93740, the ‡Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin 53706, and the **Department of Life Science and Institute for Natural Sciences, Hanyang University, Seoul 133-791, Republic of Korea

Background: Thr127 and Ser128 are important for the function of the Escherichia coli CRP (cAMP receptor protein).

Results: Thr/Ser, Thr/Thr, and Thr/Asn pairs at the positions are optimal for CRP function.

Conclusion: Evolutionarily conserved residue pairs at the positions provide high cAMP affinity while keeping CRP inactive without cAMP.

Significance: There are multiple evolutionary strategies for cAMP sensing in CRP.

The Escherichia coli cAMP receptor protein (CRP) requires cAMP binding to undergo a conformational change for DNA binding and transcriptional regulation. Two CRP residues, Thr127 and Ser128, are known to play important roles in cAMP binding through hydrogen bonding and in the cAMP-induced conformational change, but the connection between the two is not completely clear. Here, we simultaneously randomized the codons for these two residues and selected CRP mutants displaying high CRP activity in a CAMP-producing E. coli. Many different CRP mutants satisfied the screening condition for high CRP activity, including those that cannot form any hydrogen bonds with the incoming CAMP at the two positions. In vitro DNA-binding analysis confirmed that these selected CRP mutants indeed display high CRP activity in response to CAMP. These results indicate that the hydrogen bonding ability of the Thr127 and Ser128 residues is not critical for the CAMP-induced CRP activation. However, the hydrogen bonding ability of Thr127 and Ser128 was found to be important in attaining high CAMP affinity. Computational analysis revealed that most natural CAMP-sensing CRP homologs have Thr/Ser, Thr/Thr, or Thr/Asn at positions 127 and 128. All of these pairs are excellent hydrogen bonding partners and they do not elevate CRP activity in the absence of CAMP. Taken together, our analyses suggest that CRP evolved to have hydrogen bonding residues at the CAMP pocket residues 127 and 128 for performing dual functions: preserving high CAMP affinity and keeping CRP inactive in the absence of CAMP.

The Escherichia coli CAMP receptor protein (CRP) regulates many genes required for catabolism of various carbon sources other than glucose (1, 2). When glucose is absent, CAMP is produced by a membrane-bound adenylate cyclase (3). CRP then binds to the CAMP molecule and undergoes a global conformational change, leading to DNA binding and transcriptional regulation (4–6). The CAMP pocket residues play pivotal roles in this CRP activation. At least six residues (Glu72, Arg77, Arg82, Thr127, Ser128) are involved in direct interactions with the incoming CAMP (7). Biochemical and mutagenic studies indicate that these CAMP pocket residues can be classified into two groups. Glu72, Arg82, Ser128, and Arg123 have a primary role in high affinity CAMP binding and Thr127 and Ser128 serve as a pivot point for the CAMP-induced conformational change (4–6, 8–10). In other words, the first group provides anchor interactions and the second provides driver interactions for the allosteric activation of CRP (11). This view is also consistent with the observation that, whereas the first four residues are universally conserved in CAMP-sensing proteins including eukaryotic ones, the Thr127 and Ser128 of CRP are only found in CAMP and its prokaryotic homologs (12).

The crystal structures of the active CRP form have been reported several times since 1981 (7, 13, 14) (an active CRP structure complexed with bound DNA is shown in Fig. 1A). However, the structures of the inactive CRP form have been solved only recently, using both crystallography and NMR spectroscopy (5, 6). It is important to note that the crystal and NMR structures for the inactive form are not identical. Most notably, in the crystal structure, the DNA-contacting F-helices are buried inside the protein, precluding any interactions with DNA (Fig. 1B). In the NMR structure, the F-helices are exposed to the solvent, but their topology in this inactive form is unable to bind DNA (Fig. 1C). In addition, in the crystal structure, two C-helix residues (Val131 and Gly132) are not resolved. Despite these differences, there is an important commonality between these two inactive structures. In both structures, the 11-residue segment of the C-helix (Val126-Phe136) enclosing the Thr127 and Ser128 residues is unstructured, unlike the active form, suggesting a much more flexible nature of the C-helix in the inactive form (Fig. 1D). In the active CAMP-bound form, this region assumes a helical structure, which gives the C-helix an exten...
Evolution of cAMP Pocket in CRP

FIGURE 1. Comparison of the active and inactive CRP structures reveals critical conformational changes upon cAMP binding. A, CRP active structure (PDB code 2CGP). DNA-contacting F-helices are shown in red, and bound cAMP molecules near the C-helices are shown as ball-and-stick models. B, crystal structure of the inactive CRP form (PDB code 3HIF). C-helices (red) are buried such that they are unable to interact with DNA. C, NMR structure of the inactive CRP form (PDB code 2WC2). F-helices (red) remain solvent exposed but undergo a 90-degree rotation. D, cAMP binding induces a coil-to-helix transition in the C-helix region. In the inactive CRP structures, the C-helices unwind (blue, NMR structure, PDB code 2WC2) or are disordered (salmon, x-ray structure, PDB code 3HIF). cAMP binding results in a complete helix transition to the active CRP conformation (green, PDB code 2CGP).

Evolution of cAMP Pocket in CRP

FIGURE 2. The MWC model of CRP and constitutively active CRP (CRP*) mutants in relationship to their cAMP affinity. A, the MWC model assumes that CRP exists as an equilibrium of two conformations (active and inactive). The ligand cAMP (indicated by the orange square) selectively binds to the pre-existing active CRP population and accumulates the cAMP-bound CRP form, thus shifting the equilibrium of CRP toward the active conformation. B, the protein equilibrium of wild type CRP is poised toward the inactive form that binds negligibly to the ligand cAMP. The population ratio of active to inactive protein \( \approx [P]_A/[P]_I \) in CRP in the absence of cAMP was estimated to be \( 1.1 \times 10^{-6} \) (10). C, the protein equilibrium of CRP* is shifted toward the active form from in reference to wild type CRP. This will increase the pre-existing active CRP conformation that is selected by the ligand cAMP for binding. Thus, CRP* will have an increase in apparent cAMP affinity, although the intrinsic affinity of its active conformation to cAMP is invariant.

The ligand cAMP (indicated by the orange square) selectively binds to the pre-existing active CRP population and accumulates the cAMP-bound CRP form, thus shifting the equilibrium of CRP toward the active conformation. The protein equilibrium of wild type CRP is poised toward the inactive form that binds negligibly to the ligand cAMP. The population ratio of active to inactive protein \( \approx [P]_A/[P]_I \) in CRP in the absence of cAMP was estimated to be \( 1.1 \times 10^{-6} \) (10). The protein equilibrium of CRP* is shifted toward the active form from in reference to wild type CRP. This will increase the pre-existing active CRP conformation that is selected by the ligand cAMP for binding. Thus, CRP* will have an increase in apparent cAMP affinity, although the intrinsic affinity of its active conformation to cAMP is invariant.

High background CRP activity in the absence of cAMP results in unnecessary transcription and therefore CRP activity should be kept low when cAMP is absent (that is, when glucose is present). Thus, the strategy of keeping CRP inactive at basal levels of cAMP should provide an evolutionary advantage. This posits that there will be some residues (interactions) serving this role in CRP. Currently, our understanding of CRP function...
has been heavily based on the study of the active form. Not surprisingly, the Thr\textsuperscript{27} and Ser\textsuperscript{28} residues have been mostly studied for their impacts on the active form. Here, we show through directed evolution and computational analysis that the dyad Thr\textsuperscript{27} and Ser\textsuperscript{28} have an additional role in keeping CRP inactive in the absence of cAMP, in addition to attaining high cAMP affinity via a hydrogen bonding network.

### Experimental Procedures

**Strains, Plasmids, and Recombinant DNA Techniques**—Strains and plasmids used for this study are summarized in Table 1. HYC7 was used as the background strain for screening active CRP mutants in vivo in the presence of cAMP and HYC4 was used as the background strain for protein purification. Oligonucleotides used for PCR were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). DNA sequencing was performed through the service of the University of Wisconsin-Madison Biotechnology Center.

**Codon Randomization and in Vivo Screening for Active CRP Mutants in the Presence of cAMP**—The codons corresponding to E. coli CRP residues 127 and 128 were simultaneously randomized using the Stratagene QuikChange\textsuperscript{®} method with pHYC11 as template (Table 1). The randomized plasmid pool was then introduced into HYC7 via transformation (Table 1). Note that the HYC7 strain lacks the crp gene, but retains cya, so it is capable of producing cAMP, which allows for the screening of active CRP mutants at cellular cAMP levels. HYC7 cells containing the CRP mutant pool were screened on Luria-Bertani (LB) plates in the presence of X-gal (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside). The blue colony color indicates the presence of CRP activity. Plasmids were isolated from these blue colonies and the crp regions were sequenced to reveal causative mutations.

**Protein Purification**—Wild type CRP and CRP mutants were all His\textsubscript{6} tagged at their C termini and therefore, they were all purified via nickel affinity chromatography using an iminodiacetic acid-based His-Bind resin (Novagen). HYC4 (crp\textsuperscript{-} cya\textsuperscript{-}) was used to overexpress these CRP proteins for purification, and consequently all the proteins were purified as apoprotein. All protein preparations were >95% homogeneous based on SDS-PAGE.

**In Vitro DNA-binding Assay Using Fluorescence Anisotropy**—DNA-binding activities of CRP mutants were measured using the fluorescence anisotropy method. A Beacon 2000 fluorescence polarization detector (Invitrogen Corp.) was used. Probe DNA was a 26-mer CRP consensus sequence (5'-GTAAATGTTGATGTACATCAGATGGAT-3'). Only one of the strands was labeled with Texas Red, as described previously (4). The DNA-binding assay was performed in 50 mM Tris-HCl (pH 8.0), 50 mM KCl, and 1 mM EDTA with a probe concentration of 5 nM in the presence of 6.4 \( \mu \)M salmon sperm DNA (nonspecific competitor).

**Monitoring the cAMP Affinity of CRP Mutants**—cAMP affinity of CRP mutants were monitored by measuring the cAMP ability to promote the DNA-binding activity of the CRP proteins. This method was well described and successfully used for CRP and Vfr (10, 23).

### Results

**Global Sequence Alignment for Searching cAMP-sensing CRP Homologs**—A global sequence alignment tool was developed based on the Needleman-Wunsch algorithm (24) and used for searching cAMP-sensing CRP homologs. In the tool, we implemented the scoring scheme with BLOSUM62 scoring matrix and Affine gap scoring by setting gap-open and extension penalties as \(-11\) and \(-1\), respectively, as used in NCBI BLASTp as default values. The searching database is nr (accessed in March 2015) and the homology of query sequence (NP_417816.1 for the E. coli CRP) was checked against all of the sequences in nr. 1,514 non-redundant sequences were found to have qualified alignment scores, i.e. higher than the score between sequences of the E. coli CRP and Synechocystis sp. PCC 6803 CRP.

**Various CRP Mutants Altered at Positions 127 and 128 Are Active in Vivo in a cAMP-producing E. coli**—A total of 400 amino acid combinations are possible for the two CRP 127 and 128 positions. To know which amino acid pairs provide high CRP activity in the presence of cAMP, we completely randomized codons for the two residues and screened the resultant mutant pool in a cAMP-producing CRP reporter strain, HYC7 (Table 1). We used this cAMP-producing E. coli strain rather than adding cAMP to HYC4 (a cAMP-deficient CRP reporter strain) because (i) HYC7 allows us to isolate active CRP mutants responding to physiological levels of cAMP and (ii) the passage of externally added cAMP through the cell membrane is not very effective (25). The crp genes in the mutant pool are in pEXT20, and therefore their expression could be controlled by isopropyl b-D-1-thiogalactopyranoside (IPTG). Initial screening of the mutant pool using IPTG concentrations higher than 1 \( \mu \)M resulted in a high proportion of blue (transcriptionally active) colonies, suggesting that many combinations generated active CRP phenotype in the cAMP-producing E. coli strain. Then, we screened the same mutant pool without using IPTG (which would only allow the basal level of protein production from the plasmid vector), and found that this “no IPTG” condition produced a substantial number of blue (active) colonies for pattern identification. Therefore, this no IPTG condition was used for final screening. Of ~6,000 colonies total, about 2.5% (corresponding to 10 amino acid combinations (~93 combinations in terms of codon) at the two positions) showed a transcriptionally active blue colony color. DNA sequencing was performed for the plasmids isolated from these blue colonies. The resultant non-identical active CRP mutants in the presence of cAMP are listed in Table 2. After confirming the causative CRP mutations, we re-transformed the plasmids back into...
The current screening condition used the same exact E. coli strain (HYC7) and observed the same transcriptional activity of CRP in a cAMP-producing CRP reporter strain (HYC7). The listed CRP mutants are non-identical in their codons.

| CRP Sequence | CRP | Amino acid (codon) at 127 | Amino acid (codon) at 128 |
|--------------|-----|--------------------------|--------------------------|
| Wild type    | Thr (ACC) | Ser (TCA) |
| Wild type    | Thr (ACC) | Ser (TCG) |
| Wild type    | Ala (GCT) | Ser (TCG) |
| T127A        | Ala (GCT) | Ser (AGC) |
| T127A        | Ala (GCA) | Ser (AGC) |
| T127C        | Cys (TGT) | Ser (TCA) |
| T127S        | Ser (TCA) | Ser (TCT) |
| T127M        | Met (ATG) | Ser (TCT) |
| S128T        | Thr (ACC) | Thr (ACA) |
| S128T        | Thr (ACA) | Thr (ACT) |
| T127L/S128N  | Leu (CTA) | Asn (AAC) |
| T127L/S128N  | Leu (TTG) | Asn (AAT) |
| T127V/S128T  | Val (GTG) | Thr (ACC) |
| T127V/S128I  | Cys (TGC) | Ile (ATC) |
| T127V/S128V  | Val (GTA) | Val (GTA) |
| T127V/S128A  | Val (GTA) | Ala (GCC) |
| T127A/S128V  | Ala (GCA) | Val (GTA) |
| T127T/S128N  | Thr (ACC) | Asn (AAC) |
| Residue frequency | 5 Thr | 4 Ala | 3 Val | 2 Cys | 2 Leu | 1 Met | 1 Ser |
|               | 8 Ser | 3 Thr | 3 Asn | 2 Val | 1 Ile | 1 Ala |

The pattern clearly shows that amino acids that potentially disrupt a coiled-coil structure such as glycine, proline, aromatic amino acids, and charged amino acids are not allowed if CRP activity is to be preserved in the presence of cAMP. This suggests that maintaining a helical structure in the region is critical for CRP function. Besides these amino acids, various substitutions in terms of size and hydrophobicity are allowed at positions 127 and 128 for retaining CRP activity in the presence of cAMP. At position 127 it seems that any amino acid residue is acceptable as long as it is neither large nor bulky. At position 128, the pattern is similar to position 127, but small-sized amino acids were found less frequently. Generally, it appears that amino acids at both positions promote CRP activity, as long as they do not sterically hinder the incoming cAMP. Additionally, this would imply that precise topology of the two residues is probably not required for CRP function. Furthermore, some hydrophobic amino acids that cannot form hydrogen bonding interactions with cAMP were found to be acceptable at both positions. It is reported that hydrophobic amino acids at these positions often afford cAMP-free activity to CRP (4). However, T127L/S128I CRP, which showed the highest constitutive activity at these positions, was not found in our screening. In fact, this T127L/S128I CRP mutant was found to be slightly inferior to the ones listed in Table 2. Although the CRP mutants listed in Table 2 were blue on the screening plate, T127L/S128I CRP was light blue, along with other CRP mutants such as T127I/S128I, T127V/S128I, and T127L/S128V. However, the same T127L/S128I CRP mutant was exclusively screened from the same 127/128 randomization mutant pool for an observable constitutive activity in a cAMP-deficient E. coli reporter strain under the same no IPTG condition (4). The current screening condition used the same exact E. coli reporter strain, except it has the cya gene in the chromosome. It is unlikely that T127L/S128I CRP has particularly poor protein expression or stability due to the presence of the chromosomal cya gene in our current screening procedure. Therefore, our failure to select T127L/S128I in our current screening is probably not due to exceptionally poor protein accumulation of this mutant inside E. coli cells. Conversely, we hypothesize that the selected hydrophobic CRP mutants (at the two positions) listed in Table 2 are not selected because of their potential constitutive activity, but because of their cAMP-dependent activity. This observation apparently contradicts the previously held notion that the hydrogen bonding ability of Thr^{127} and Ser^{128} is critical for the cAMP responsiveness of CRP.

The in Vivo-selected CRP Mutants at Positions 127 and 128 Show cAMP-dependent DNA Binding — The in vivo results above are informative, but they are affected by a variety of complicating factors, such as protein level and interaction with RNA polymerase, and therefore fail to provide direct information about cAMP responses of the CRP mutants. Therefore, some of the selected CRP mutants were purified for the measurement of in vitro DNA-binding activity; all of the tested CRP mutants bound DNA in a cAMP-dependent manner (Table 3). They did not show any significant DNA binding in the absence of cAMP. In the presence of 0.1 mM cAMP, these CRP mutants exhibited high DNA-binding activity. Their K_d values for DNA narrowly ranged from 11 to 35 nM, suggesting that these CRP mutants are very similar in their DNA-binding affinity when they are saturated with cAMP. Notably, some of the CRP mutants are absolutely defective in hydrogen bonding at both positions 127 and 128 (e.g. T127V/S128V, T127V/S128A, and T127A/S128V). The results demonstrate that (i) Thr^{127} and Ser^{128} are not critical for the cAMP-induced conformational change and (ii) hydrogen bonding interactions between the residues of 127, 128, and cAMP are not necessary for the change.

The Selected CRP Mutants at Positions 127 and 128 Have Different cAMP Affinity — We then measured the cAMP affinity in vitro for the 13 purified CRP mutants altered at positions 127 and 128. For the assessment of cAMP affinity, we used a functional approach of measuring effective cAMP concentrations that stimulate DNA binding for each CRP protein. This coupled cAMP affinity assay has already been established for CRP and Vfr in analyzing their ligand affinity and binding modes (10, 23). We used this coupled method because it provides physiologi-
Evolution of cAMP Pocket in CRP

Certain Amino Acid Substitutions at Positions 127 and 128 Increase cAMP Affinity through Constitutive Activity, Which Is Independent of Hydrogen Bonding—In the absence of cAMP, the CRP Thr\textsuperscript{127} and Ser\textsuperscript{128} residues create an imperfect coiled-coil interaction at the dimer interface, thereby ensuring that it is the ligand cAMP that activates the protein (4–6). For this reason, substitutions at the two positions that strengthen the coiled-coil interaction confer cAMP-free, constitutive activity to CRP (4). Separately, we have shown that the presence of constitutive activity leads to increased cAMP affinity in CRP through the equilibrium-shift mechanism, independent of hydrogen bonding (10). These considerations predict that CRP 127 and 128 residues exert their impacts on cAMP affinity in two ways: (i) via their ability to form hydrogen bonds and (ii) by their ability to promote constitutive activity. Then, what will happen to cAMP affinity if a CRP mutant at positions 127 and 128 gains constitutive activity, but loses hydrogen bonding ability? This is relevant because many of the selected CRP mutants lost hydrogen bonding ability at one or both positions (Table 2). Although their constitutive activity was undetectable up to 1 μM protein concentration in our in vitro assay, their protein equilibria still might have been shifted toward the active form. This is because wild type CRP is extremely shifted toward the inactive form (in the absence of cAMP) (10) so much that a significant shift toward the active form may not be detectable.

In an effort to directly monitor the interplay between constitutive activity and hydrogen bonding in modulating cAMP affinity, we created a CRP mutant (T127L/S128V) altered at the two positions and its two parental single mutants (T127L and S128V), and compared their cAMP-free DNA-binding activity and cAMP affinity. As shown in Fig. 4A, the T127L/S128V CRP mutant demonstrated detectable constitutive (cAMP-free) activity. However, the two parental single CRP mutants T127L and S128V did not display any observable cAMP-free activity (Fig. 4A). On the other hand, the T127L/S128V CRP mutant lacks hydrogen bonding at both positions, whereas the two single CRP mutants retain one hydrogen bonding residue at either position. Therefore it can be concluded that compared with the two single CRP mutants, the T127L/S128V CRP mutant significantly gained constitutive activity, but lost hydrogen bonding interactions. Then, we compared their cAMP affinity. As shown in Fig. 4B, T127L/S128V CRP actually displayed a higher cAMP affinity than T127L or S128V CRP. Our interpretation for this phenomenon is that the gain of cAMP affinity due to the increase of constitutive activity was more significant than the decrease in cAMP affinity due to the loss of the second hydrogen bond. Thus, this result supports the notion that CRP mutants at positions 127 and 128 have differential cAMP affinity due to combined effects of hydrogen bonding ability and protein equilibrium poise.

Our analysis also predicts that CRP mutants at positions 127 and 128 can be classified into four categories: 1) low cAMP affinity and low cAMP-free activity, 2) high cAMP affinity and high cAMP-free activity, 3) high cAMP affinity and low cAMP-free activity, and 4) low cAMP affinity and high cAMP-free activity. The data presented below show that most of the cAMP-sensing CRP homologs found in bacteria have the phenotype of category 3.

The last tier mutants include T127V/S128V CRP (VV CRP), T127V/S128A CRP (VA CRP), T127A/S128V CRP (AV CRP), and T127S CRP (SS CRP) along with the above mentioned MS and VT CRP. The VV, VA, and AV CRP lack hydrogen bonding ability at both positions. This again shows the importance of the hydrogen bonding ability for cAMP affinity. On the other hand, the SS, MS, and VT CRP have hydrogen bonding ability, so there must be another factor(s) in these mutants that lower their functional cAMP binding.

FIGURE 3. CRP mutants at positions 127 and 128 differ in their cAMP affinity. cAMP affinity was assessed by the requirement of CRP proteins for DNA binding. For each, the reaction solution containing 200 nM protein was titrated by various concentrations of cAMP. The in vitro DNA binding was monitored using a fluorescence anisotropy method. DNA-binding activities of these mutants without cAMP are similar to the activities at their lowest cAMP concentrations within experimental errors.
Thr/Ser, Thr/Thr, and Thr/Asn Are Predominant Residue Pairs at the Positions Analogous to CRP Thr127 and Ser128 in cAMP-sensing CRP Homologs—We then comprehensively surveyed the distribution of residue pairs at the analogous positions to CRP Thr127 and Ser128 in all sequentially identified CRP homologs that are predicted to sense cAMP. For this, we searched the entire NCBI nr database using the global sequence alignment procedure (as detailed under “Experimental Procedures”) with the E. coli CRP as the query. First we defined a CRP homolog to be cAMP sensing if it meets the following two conditions. (i) The homology score of the CRP homolog with the E. coli CRP is better than the one between the E. coli CRP and Synechocystis sp. PCC 6803 homolog (BAA18260.1). The PCC 6803 CRP (27) was chosen because it is the most sequentially distant CRP homolog from the E. coli CRP among the experimentally proven cAMP-sensing homologs. According to our global alignment scheme, the PCC 6803 CRP was more distant than CRP proteins from Corynebacterium glutamicum (28), Thermus thermophilus HB8 (29), Mycobacterium tuberculosis H37Rv (30), and Streptomyces coelicolor (31). (ii) The CRP homolog retains all the amino acids that are critical in the E. coli CRP for high affinity cAMP binding: Glu at position 72, Arg at position 82, either Ser or Thr at position 83, and Arg at position 123.

A total of 1,514 cAMP-sensing CRP homologs from the NCBI nr database met the first condition (i) (supplemental Table S1). Then, 25 sequences were manually eliminated from analysis: 22 were from engineered or mutated E. coli CRP proteins and three lacked an amino acid (a gap) at the position analogous to E. coli Thr127 in their alignment. The resultant 1,489 sequences were further analyzed. A total of 29 different combinations of residue pairs at positions 127 and 128 were found (Table 5). Generally, these combinations are reminiscent of the experimentally selected combinations (Table 2). However, three residue pairs (Thr/Ser, Thr/Thr, and Thr/Asn) were found with high frequency (80% combined), suggesting that the three residue pairs are particularly good at serving CRP function at these positions and therefore were evolutionarily selected against other amino acid combinations. Importantly, all of these residue pairs provide high cAMP affinity when placed in the E. coli CRP. Furthermore, these residue pairs do not cause cAMP-free activity (Table 3) and do not resemble the residues that stimulate cAMP-free activity (4). These considerations led us to propose that Thr/Ser, Thr/Thr, and Thr/Asn residue pairs appear with high frequency (among cAMP-sensing CRP homologs) because they fulfill dual functions: providing high cAMP affinity and keeping CRP inactive in the absence of cAMP.

In contrast, the other 26 residue pairs (led by Ala/Asp, Ala/Asn, Val/Asn, and Leu/Thr pairs in occurrence) are compromised in their hydrogen bonding ability in at least one position. Our data suggest that these minor residue pairs would result in lower cAMP affinity in the context of the E. coli CRP. However, it does not necessarily mean that the CRP homologs containing these residue pairs would have lower cAMP affinity than does the E. coli CRP. This is because the residue pair is only one of the factors affecting cAMP affinity. Characterizing the cAMP responsiveness of these CRP homologs is required to understand the role of these residue pairs in the context of their native proteins. Last, according to our in vivo selection and in vitro
TABLE 5
Distribution of residue pairs at positions 127 and 128 in all sequentially-identified cAMP-sensing CRP homologs found in the NCBI nr database

| Residue 127 | Residue 128 | Occurrence |
|-------------|-------------|------------|
| Ala         | Asp         | 68         |
| Gly         | Asp         | 1          |
| Leu         | Asp         | 2          |
| Met         | Asp         | 2          |
| Ala         | Gln         | 1          |
| Leu         | Gln         | 1          |
| Met         | Gln         | 1          |
| Ile         | Ile         | 1          |

Discussion

CRP residues Thr$^{127}$ and Ser$^{128}$ have been considered the epicenter for the propagation of the conformational change of the protein required for DNA binding (4–6). Popovych et al. (6) hypothesized that direct interactions between Thr$^{127}$/Ser$^{128}$ and the cAMP adenine base are critical for the C-helix coil-to-helix transition accompanying the allosteric transition of CRP. However, as shown in Table 3, many substitutions at positions 127 and 128, including those that cannot make direct hydrogen bonding interactions with cAMP, retain wild type-level DNA binding. Thus, our results disprove the hypothesis that such direct interaction is the primary contributor to the critical coil-to-helix transition. Rather, our data suggest that most CRP interactions leading to activity saturation are different and span roughly 2 orders of magnitude (1–100 μM) (Fig. 3). Because these CRP mutants are all active in vivo, the measured range of cAMP affinity may reflect the intracellular cAMP levels of E. coli or vice versa. A comprehensive study reported that the intracellular cAMP levels of E. coli fluctuate from 0.5 to 8 pmol/10^8 cells in response to the variation of external glucose levels (32). The volume of an E. coli cell is ~1.0 × 10^{-15} liters (33), so cAMP levels vary from 5 to 80 μM in concentration. However, these values should be taken with the following consideration. It was reported that intracellular cAMP levels in bacteria are very difficult to accurately determine (34). In fact, our data suggest that this experimentally determined lower limit of cAMP levels (5 μM) may be an overestimate. This is because 5 μM would activate many of the CRP mutants for DNA binding (shown in Fig. 3) when they should remain inactivated, resulting in unwanted transcriptional activation in E. coli cells.

Evolution of cAMP Pocket in CRP

DNA-binding/cAMP affinity analyses, the Cys/Ser pair at positions 127 and 128 should not be much different from the predominant trio (Thr/Ser, Thr/Thr, and Thr/Asn), however, the Cys/Ser pair has never been found in natural cAMP-sensing CRP homologs. Currently, we do not know the basis for the complete absence of this pair. We can only speculate that Cys at position 127 provides some deleterious effect on CRP in some way of which we are currently unaware.
affinity can be elevated in a CRP mutant at these positions: hydrogen bonding and protein equilibrium shift toward the active form. These two forces can be combined in several ways to generate a greater number of residue combinations at the two positions that meet the threshold for cAMP affinity. For example, both residues can be good at hydrogen bonding, one residue is good at hydrogen bonding and the other is good at shifting protein equilibrium toward the active form, or both residues are good at shifting the protein equilibrium toward the active form. The best example for the last case is the constitutively active T127L/S128I CRP mutant, which lacks hydrogen bonding ability at both positions, but possesses an extremely shifted protein equilibrium poise toward the active form (4). Consequently, this mutant continues to bind cAMP with observable affinity for further activation of the protein (4). Nevertheless, this T127L/S128I CRP mutant ranked below the selected CRP mutants in our in vivo selection, which probably indicates that its cAMP affinity is below the threshold level. This may suggest that the hydrogen bonding ability at CRP 127 and 128 positions is more important than their equilibrium shift ability in attaining physiologically relevant high cAMP affinity.

In cAMP-sensing CRP homologs, three residue pairs (Thr/ Ser, Thr/Thr, and Thr/Asn) are predominantly found at the positions analogous to E. coli residues Thr^{127} and Ser^{128}. The Thr/Ser pair is found in CRP from E. coli and closely related species such as Salmonella enterica (35) and Vibrio fischeri (36). The Thr/Thr pair is found in Vfr from Pseudomonas aeruginosa (37) and related species (38). The Thr/Asn pair is found in CRP from M. tuberculosis (26, 30) and related species, which include C. glutamicum (28) and S. coelicolor (31). Several of these CRP proteins have been crystallized, in which the dyads are visualized to interact with cAMP through hydrogen bonding interactions (7, 28, 30, 37). All of the naturally occurring dyads exhibit a comparable hydrogen bonding pattern to the adenosine ring of cAMP (Fig. 6). In all cases, the residue at the second position (Ser, Thr, or Asn) accepts a hydrogen bond from the N6 of cAMP and donates a hydrogen bond to the N7 of cAMP. At the first position, all of the naturally occurring dyads contain a Thr residue (Thr^{127} in the E. coli CRP) that accepts a hydrogen bond from the N6 of cAMP (Fig. 6, A–C). Interestingly, the Cys^{127} substitution found in our directed evolution of the E. coli CRP appears to be functionally equivalent to the Thr residue, accepting a hydrogen bond from the N6 of cAMP (Fig. 6D). All of these pairs, including Cys/Ser, provide high cAMP affinity (Fig. 3). Thus, there is a correlation between hydrogen bonding and high cAMP affinity. This result suggests that hydrogen bonding ability at positions 127 and 128 is important for high cAMP affinity.

Although the above three residue pairs at positions 127 and 128 are roughly equivalent in the context of the E. coli CRP (in terms of hydrogen bonding with cAMP and cAMP affinity), this does not necessarily mean that naturally occurring CRP homologs having these residue pairs will have similar cAMP affinity. The P. aeruginosa Vfr and Pseudomonas putida CRP (containing the Thr/Thr pair) are known to exhibit much higher cAMP affinity than the E. coli CRP (23, 38). In the P. aeruginosa Vfr, it is hypothesized that “Thr” at the second position, compared with the CRP residue “Ser,” shifts the protein equilibrium toward the active form while maintaining a similar hydrogen bonding ability, thereby resulting in higher cAMP affinity (23). Consequently, T133S Vfr, which has Ser instead of Thr at the analogous position, showed dramatically reduced cAMP affinity (23). Nevertheless, in the E. coli CRP context, the effect of Thr substitution at the second position is only marginal in best terms of increasing cAMP affinity (Fig. 3). This suggests that the effect of Thr at the second position on cAMP affinity is protein context-dependent. The E. coli CRP may have some other factor(s) that make Ser at the second position exceptionally good in terms of cAMP affinity. Alternatively, the P. aeruginosa Vfr may have factor(s) that make Ser at the second position relatively poor in terms of cAMP affinity. In contrast, M. tuberculosis CRP (containing the Thr/Asn pair) is known to have low cAMP affinity (39). Based on our results, this relatively low cAMP affinity displayed by the M. tuberculosis CRP cannot originate from the presence of Asn in replacement of Ser at the second position. The M. tuberculosis CRP has also conserved CAMP pocket residues other than the Asn (26), thus, there is no reason to believe that the hydrogen bonding efficiency of the protein with cAMP is compromised. We suppose that the basis for this low cAMP affinity for the M. tuberculosis CRP must be a shifted protein equilibrium poise toward the inactive form. Further study is required to prove or disprove this hypothesis.

Currently, little is known about the CRP homologs having uncommon 26 residue pairs other than Thr/Ser, Thr/Thr, and Thr/Asn at the positions analogous to the E. coli CRP 127 and 128. Characterization of these CRP homologs containing these minor residue pairs will help understand the impact of these pairs on cAMP affinity in their natural protein backgrounds. The characterization will also reveal the range of cAMP affinity (upper and lower boundaries) for a cAMP-sensing CRP homolog in bacteria.

Finally, what might be the required functions that need to be fulfilled by the CRP 127 and 128 residues? Our directed evolu-
tion, cAMP affinity measurement, and computational analysis suggest that CRP 127 and 128 residues have evolved to activate CRP with low amounts of cAMP. To attain this property, CRP has evolved to build an elaborate hydrogen bonding network at positions 127 and 128. Although both hydrogen bonding and protein equilibrium shift can be used for gaining high cAMP affinity, only hydrogen bonding does not elevate cAMP-free activity, thereby ensuring that CRP activity continues to be controlled by the ligand cAMP. The three most common residue pairs (Thr/Ser, Thr/Thr, and Thr/Asn) are all excellent hydrogen bonding formers. Their conservation at the two cAMP-contacting positions in cAMP-sensing CRP homologs is the direct outcome of an evolutionary strategy for providing two dissimilar functions: high cAMP affinity and low cAMP-free CRP activity.

Author Contributions—H. Y., S. M. G., M. N. H., and C. V. S. designed the study and wrote the paper. J. P. performed the global alignment analysis. H. Y. and C. V. S. interpreted the global alignment analysis. H. Y. and C. V. S. performed the in vivo screening and analysis. J. S. purified the CRP proteins. H. Y., J. J. G., and J. W. L. performed the in vitro DNA-binding analysis and cAMP affinity analysis. C. L. B. performed the structural analysis. All authors analyzed the results and approved the final version of the manuscript.

References

1. Kolb, A., Busby, S., Buc, H., Gargouri, S., and Adhya, S. (1994) Transcriptional regulation by cAMP and its receptor protein. Annu. Rev. Biochem. 62, 749–795.
2. Busby, S., and Ebright, R. H. (1999) Transcription activation by catabolite activator protein (CAP). J. Mol. Biol. 293, 199–213.
3. Holgerson, B. M., Arents, J. C., Bader, E., Eijkemans, K., Inada, T., Aiba, H., and Postma, P. W. (1998) Inducer exclusion by glucose 6-phosphate in Escherichia coli. Mol. Microbiol. 28, 755–765.
4. Youn, H., Kerby, R. L., Conrad, M., and Roberts, G. P. (2006) Study of highly constitutively active mutants suggests how cAMP activates cAMP receptor protein. J. Biol. Chem. 281, 1119–1127.
5. Sharma, H., Yu, S., Kong, J., Wang, J., and Steitz, T. A. (2009) Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding. Proc. Natl. Acad. Sci. U.S.A. 106, 16604–16609.
6. Popovych, N., Tzeng, S. R., Tonelli, M., Ebright, R. H., and Kalodimos, C. G. (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. Proc. Natl. Acad. Sci. U.S.A. 106, 6927–6932.
7. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science 253, 1001–1007.
8. Ebright, R. H., Le Grice, S. F., 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.
9. Youn, H., Kerby, R. L., Koh, J., and Roberts, G. P. (2007) A C-helix residue, Arg-123, has important roles in both the active and inactive forms of the cAMP receptor protein. J. Biol. Chem. 282, 3632–3639.
10. Youn, H., Koh, J., and Roberts, G. P. (2008) Two-state allosteric modeling suggests protein equilibrium as an integral component for cyclic AMP (cAMP) specificity in the cAMP receptor protein of Escherichia coli. J. Bacteriol. 190, 4532–4540.
11. Nussinov, R., and Tsi, C. I. (2014) Unraveling structural mechanisms of allosteric drug action. Trends Pharmacol. Sci. 35, 256–264.
12. Berman, H. M., Ten Eyck, L. F., Golds, D. S., Haste, N. M., Kornev, A., and Taylor, S. S. (2005) The cAMP binding domain: an ancient signaling module. Proc. Natl. Acad. Sci. U.S.A. 102, 45–50.
13. McKay, D. B., and Steitz, T. A. (1981) Structure of catabolite gene activator protein at 2.9-Å resolution suggests binding to left-handed B-DNA. Nature 290, 744–749.
14. Passner, J. M., Schultz, 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.
15. Gronenborn, A. M., Sandulache, R., Gartner, S., and Clore, G. M. (1988) Mutagenesis of the cyclic AMP binding site of the cyclic AMP receptor protein of Escherichia coli. Biochem. J. 253, 801–807.
16. Lee, E. J., Glassow, 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.
17. Leu, S. F., Baker, C. H., Lee, E. J., and Harman, J. G. (1999) Position 127 amino acid substitutions affect the formation of CRP:cAMP:RNA polymerase complexes but not CRP:cAMP:RNA polymerase complexes at lacP. Biochemistry 38, 6222–6230.
18. Cheng, X., Kovac, L., and Lee, J. C. (1995) Probing the mechanism of CRP activation by site-directed mutagenesis: the role of serine 128 in the allosteric pathway of cAMP receptor protein activation. Biochemistry 34, 10816–10826.
19. Harman, J. G. (2001) Allosteric regulation of the cAMP receptor protein. Biochim. Biophys. Acta 1547, 1–17.
20. Monod, J., Wyman, J., and Changeux, J. P. (1965) On the nature of allosteric transitions: a plausible model. J. Mol. Biol. 12, 88–118.
21. Kumar, S., Ma, B., Tsai, C. J., Sinha, N., and Nussinov, R. (2000) Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci. 9, 10–19.
22. Ma, B., Shatsky, M., Wolfson, H. J., and Nussinov, R. (2002) Multiple diverse ligands binding at a single protein site: a matter of preexisting populations. Protein Sci. 11, 184–197.
23. Serate, J., Roberts, G. P., Berg, O., and Youn, H. (2011) Ligand responses of Vfr, the virulence factor regulator from Pseudomonas aeruginosa. J. Bacteriol. 193, 4859–4868.
24. Needleman S. B., and Wunsch C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.
25. Rya, M. H., Youn, H., Kang, I. H., and Golensky, M. (2015) Identification of bacterial guanylate cyclases. Proteins 83, 799–804.
26. Stapleton, M., Haq, H., Hunt, D. M., Arvign, K. B., Artyomov, P. J., Buxton, R. S., and Green, J. (2010) Mycobacterium tuberculosis cAMP receptor protein (Rv3676) differs from the Escherichia coli paradigm in its cAMP binding and DNA binding properties and transcription activation properties. J. Biol. Chem. 285, 7016–7027.
27. Forcada-Nadal, A., Forchhammer, K., and Rubio, V. (2014) SPR analysis of promoter binding of Synechocystis PCC6803 transcription factors NtcA and CRP suggests cross-talk and sheds light on regulation by effector molecules. FEBS Lett. 588, 2270–2276.
28. Townsend, P. D., Jungwirth, B., Pojer, F., Bußmann, M., Money, V. A., Cole, S. T., Pühler, A., Tauch, A., Bott, M., Cann, M. J., and Pohl, E. (2014) The crystal structures of apo and cAMP-bound GlxR from Corynebacterium glutamicum reveal structural and dynamic changes upon cAMP binding in CRP:FNRI family transcription factors. PLoS ONE 9, e113265.
29. Shinaki, A., Kira, S., Nakagawa, N., Kashiura, A., Kuramitsu, S., and Yokoyama, S. (2007) Transcription activation mediated by a cyclic AMP receptor protein from Thermus thermophilus HB8. J. Bacteriol. 189, 3891–3901.
30. Reddy, M. C., Palaninathan, S. K., Bruning, J. B., Thurman, C., Smith, D., and Sacchettini, J. C. (2009) Structural insights into the mechanism of the allosteric transitions of Mycobacterium tuberculosis cAMP receptor protein. J. Biol. Chem. 284, 36581–36591.
31. Derouxaux, A., Dehaeng, D., Lecocq, E., Halici, S., Nothaf, H., Giannotta, M., Moutouzoulis, G., Duart, J., Devreese, B., Tegtemeyer, F., Van Bemuyn, J., and Rigali, S. C. 2004 Crp of Streptomyces coeliotica is the third transcription factor of the large CRP-FNR superfamily able to bind cAMP. Biochem. Biophys. Res. Commun. 325, 983–990.
32. Notley-McRobb, L., Death, A., and Ferenci, T. (1997) The relationship between external glucose concentration and cAMP levels inside Esche-
Evolution of cAMP Pocket in CRP

*richia coli*: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. *Microbiology* 143, 1909–1918

33. Phillips, R., Kondev, J., and Theriot, J. (2009) Physical Biol. Cell, Garland Science, New York

34. Bettenbrock, K., Sauter, T., Jahreis, K., Kremling, A., Lengeler, J. W., and Gilles, E. D. (2007) Correlation between growth rates, EIIACrr phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12. *J. Bacteriol.* 189, 6891–6900

35. Villarreal, J. M., Hernández-Lucas, I., Gil, F., Calderón, I. L., Calva, E., and Saavedra, C. P. (2011) cAMP receptor protein (CRP) positively regulates the yihL-yshA operon in *Salmonella enterica* serovar Typhi. *Microbiology* 157, 636–647

36. Lyell, N. L., Colton, D. M., Bose, J. L., Tumen-Velasquez, M. P., Kimbrough, J. H., and Stabb, E. V. (2013) Cyclic AMP receptor protein regulates pheromone-mediated bioluminescence at multiple levels in *Vibrio fischeri* ES114. *J. Bacteriol.* 195, 5051–5063

37. Cordes, T. J., Worzalla, G. A., Ginster, A. M., and Forest, K. T. (2011) Crystal structure of the *Pseudomonas aeruginosa* virulence factor regulator. *J. Bacteriol.* 193, 4069–4074

38. Arce-Rodríguez, A., Durante-Rodríguez, G., Platero, R., Krell, T., Calles, B., and de Lorenzo, V. (2012) The Crp regulator of *Pseudomonas putida*: evidence of an unusually high affinity for its physiological effector, cAMP. *Environ. Microbiol.* 14, 702–713

39. Green, J., Stapleton, M. R., Smith, L. J., Artymiuk, P. J., Kahramanoglou, C., Hunt, D. M., and Buxton, R. S. (2014) Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. *Curr. Opin. Microbiol.* 18, 1–7

40. Baichoo, N., and Heyduk, T. (1999) DNA-induced conformational changes in cyclic AMP receptor protein: detection and mapping by a protein footprinting technique using multiple chemical proteases. *J. Mol. Biol.* 290, 37–48

41. Dykxhoorn, D. M., St Pierre, R., and Linn, T. (1996) A set of compatible *tac* promoter expression vectors. *Gene* 177, 133–136