Cyclic AMP Receptor Protein: Role in Transcription Activation

Benoit de Crombrugghe, Stephen Busby, Henri Buc

Summary. The structure of this pleiotropic activator of gene transcription in bacteria and its interaction sites at promoter DNA's as well as the role of this protein in the RNA polymerase-promoter interactions are reviewed.

Cyclic adenosine monophosphate (AMP) serves a central role in controlling the catabolic activity of both prokaryotic and eukaroyotic cells, but its effects are mediated by two very different mechanisms in these two cell types. In bacteria, all known effects of cyclic AMP are mediated by the cyclic AMP receptor protein (CRP or CAP), whereas in eukaryotes they are mediated by a cyclic AMP–dependent protein kinase. In both types of cells the role of these proteins is to sense the intracellular concentration of cyclic AMP. In bacteria these levels vary primarily with the carbohydrate substrate on which the cells grow (*1*). When the intracellular levels of cyclic AMP increase, CRP activates the expression of a series of genes in bacteria, whereas in animal cells the cyclic AMP–dependent protein kinase responds to this signal by phosphorylating other enzymes or proteins, for example, those responsible for the breakdown of glycogen. Among the genes that are activated in bacteria in response to an increase in cyclic AMP are those that encode the enzymes for the catabolism of lactose, arabinose, maltose, and other sugars. These genes can be fully expressed only when the intracellular concentrations of cyclic AMP are high. In this article we address the question of how CRP activates gene expression in bacteria. Necessary steps in the activation process include the binding of cyclic AMP to the receptor and the interaction of this complex at the promoters of catabolite sensitive genes.

That transcription of these genes is indeed activated by the cyclic AMP–dependent binding of CRP at specific promoters was demonstrated in several ways. (i) Expression of catabolite sensitive genes is reduced in cells that harbor mutations in the genes for adenylate cyclase or for CRP (*2*); exogenously added cyclic AMP compensates for the effects of a deficient adenylate cyclase gene. (ii) Promoters of operons whose expression depends on cyclic AMP and CRP contain specific sites at which cyclic AMP–CRP can bind (*3–8*). (iii) In vitro transcription by RNA polymerase of DNA fragments containing these promoters is dependent on the presence of cyclic AMP–CRP (*9–11*). (iv) In some cases mutant promoters have been isolated at which cyclic AMP–CRP is unable to bind. At these promoters cyclic AMP–CRP fails to activate transcription both in intact cells and in vitro (*12–16*).

In this article, we review a series of recent studies, including biochemical and genetic analyses of a number of CRP binding sites, an x-ray crystallographic study of CRP and its comparison with the structure of other DNA binding regulatory proteins, an analysis of CRP binding to DNA, and a study of the step in the RNA polymerase-promoter interactions at which CRP exerts its effect to activate transcription. Although we have gained an accurate view of the DNA sequences that are recognized by the protein and have obtained a detailed picture of the three-dimensional structure of CRP, we are still lacking (i) an understanding of the precise interactions between the cyclic AMP–CRP complex and its binding sites at promoters, and (ii) a comprehension of the relation between the structure of CRP-promoter complexes and the mechanism of transcription activation.

Comparison of CRP Binding Sites

The extent to which the different CRP binding sites have been characterized varies from one promoter to another. The few promoters for which CRP binding site mutations exist and for which detailed chemical probe protection data are available obviously provide most of the information. However, a comparative analysis of sequences of the various sites that have been studied has been made (*7, 17*) and provides additional information (Figs. 1 and 2). The points that emerge from such an analysis are as follows.

1) One CRP dimer binds at each *gal*, *mal*, and *lac* promoter site (*18–20, 21*). By analogy, a similar stoichiometry may be assumed for other sites. Indeed, the size of the DNA segment protected by CRP against deoxyribonuclease attack is fairly uniform, approximately 25 base pairs (bp). The bases protected by cyclic AMP–CRP from chemical attack and the mutations that prevent CRP binding are located within these sequences (*3–5*). The various chemical protection experiments with the *lac* promoter indicate that the protein makes its major contacts in two successive major grooves of the DNA template, but the entire area of interaction spans these two successive major grooves, with the minor groove in between and one minor groove at each end (*3*).

Within the segment protected by CRP against deoxyribonuclease I, by far the most conserved sequence is 5′TGTGA3′ (T, thymine; G, guanine; A, adenine) (Fig. 2). All the evidence indicates that the TGTGA sequence is critical for CRP binding. For example, in methylation protection experiments, the protected guanines fall in this sequence in the three cases that have been examined (*3–5*). The protein recognizes the residues of this motif that appear in the major groove.

Furthermore, genetic data emphasize the importance of the sequence, 5′TGTGA3′. At the *gal* site, all three point mutations known to prevent stable CRP binding are located in this sequence (*15*); at the *ara* site, a 3-bp deletion in the 5′TGTGA3′ motif prevents CRP binding (*16*); and at the *lac* promoter, point mutations that decrease CRP binding fall in this sequence (*12–14*). To detect other homologies, the different sequences have been aligned at their 5′TGTGA3′ motif since no other alignment reveals any significant homology.

Benoit de Crombrugghe is on the staff of the Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205. Stephen Busby was and Henri Buc is at the Département de Biologie Moleculaire, Institut Pasteur, Paris, 75015 France. The present address of Stephen Busby is Department of Biochemistry, University of Birmingham, Birmingham B152TT, England.

From this analysis a "consensus" sequence emerges (Fig. 1) which consists of (i) the sequence TGTGA, which is highly conserved between different sites, and includes a GA doublet that appears in all CRP sites; (ii) a six–base pair block that shows little sequence preference; (iii) a second sequence, less well conserved, that contains an inverted repeat of the TGTGA motif.

2) In rare cases, the sequence 6-bp downstream of the TGTGA motif generates a twofold symmetry in the DNA sequence that is perfect (as in *tna*A) or nearly perfect (as in *lac*) (*10, 14*). In the case of *lac*, this second sequence, and hence the symmetry, appears to be functional. Evidence for this view comes from both genetic and biochemical data. Firstly, the two CRP site mutations that have been isolated in this site cause a transition of G · C to A · T (C, cytosine) in two loci, which are symmetrically related about the dyad axis (*14*). Furthermore, the bases protected by CRP from attack by dimethylsulfate and ultraviolet irradiation, as well as the phosphates that are necessary for CRP binding, as probed by the ethylating agent ethylnitrosourea, are also symmetrically placed around the dyad (*3*) (Fig. 1). Taken together, these results show that, when CRP is bound at the *lac* site, the sequence in both of the two symmetrically arranged elements must recognize approximately the same protein configuration. Because one dimer of CRP binds to this zone, it is likely that the subunits of the dimer are arranged in such a way that they each recognize one of the two elements (*22*).

3) In most cases, the symmetry described above is not observed. An important question, thus, is whether nonsymmetric sequences following the 5′TGTGA3′ motif are important. In *gal*, a deletion analysis showed that removal of these sequences blocked the action of CRP in vivo and in vitro. Thus, the sequence is important for CRP binding even though it is not symmetric to the

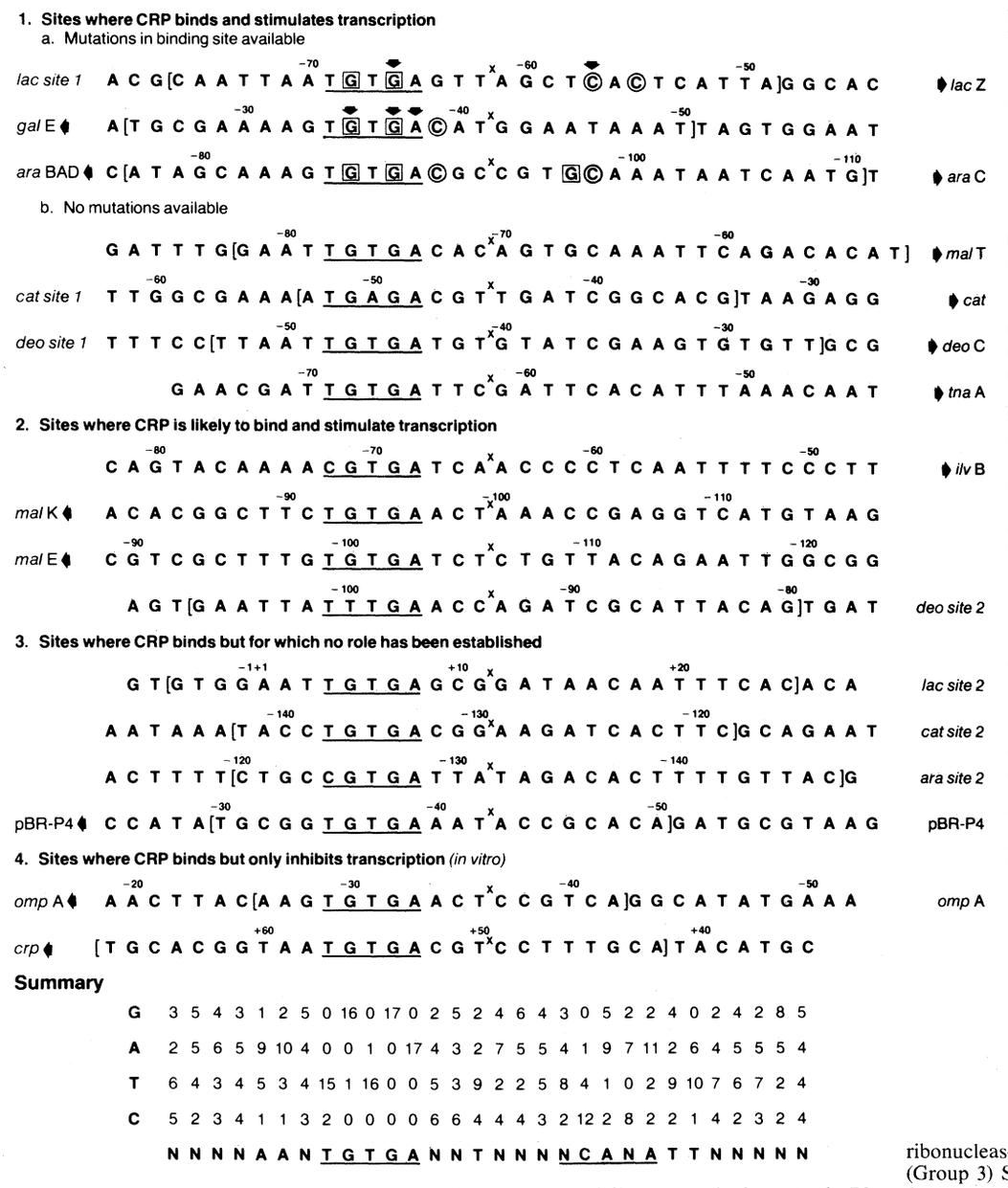

Fig. 1. Sequences of CRP binding sites. The figure shows the sequences around 17 sites specifically recognized by cyclic AMP-CRP. The sequences are aligned by the underlined motif 5′TGTGA3′, which is found in 13 out of the 17 sequences. The other four sequences are aligned by the best fit to this sequence, which is also underlined. Numbers above the sequences indicate distances in base pairs from the corresponding transcription start point. Horizontal arrows indicate the direction of transcription relative to the CRP site sequences. Next to these arrows the symbol of the first transcribed gene is indicated. In the case of *ara*, where divergent transcription occurs, the sequence is numbered with respect to the *ara*-BAD transcription start point. In each sequence, the brackets mark the outermost nucleotides protected by CRP in deoxyribonuclease protection studies. Since many studies were performed in different laboratories under different conditions, these brackets can only be regarded as approximate limits of each CRP site. The sites are classified in four groups: (Group 1) Sites where CRP binds and stimulates transcription. The position of the mutations that define the CRP binding site are indicated by vertical arrows. Guanines protected by CRP from modification with dimethylsulfate are boxed, whereas cytosines which are complementary to protected guanines are circled. (Group 2) Sites where CRP is likely to bind and stimulate transcription, but for which no ribonuclease protection data are available. (Group 3) Sites where CRP binds but for which no role has been established. This group includes the second CRP site at the *lac, cat, deo*P2, and *ara* promoters. It also includes the site at the pBR-P4 promoter where CRP stimulates transcription but only in vitro, not in intact cells. (Group 4) Sites where CRP binds but only inhibits transcription in vitro. The summary shows the frequency of appearance of each of the four bases at each position. Underneath we deduce a consensus sequence comprised of nucleotides that appear nine or more times at each position throughout the 17 cases examined.

5′TGTGA3′ motif. Since this sequence is protected by the cyclic AMP–CRP complex against deoxyribonuclease attack, at least some of its elements are recognized by the second subunit of the CRP dimer and must contribute to the overall stability of the complex (*4, 18, 23*). We propose, therefore, that the two subunits of CRP recognize two zones of sequences separated by 6 bp. One of these contains the sequence 5′TGTGA3′. The other stretch contains either a symmetrically arranged second version of this sequence or another type of sequence.

4) Affinity of CRP for DNA appears to be higher when the sequence downstream of the conserved 5′TGTGA3′ motif is symmetrical than when it is not symmetrical. The binding of the cyclic AMP–CRP complex to a fragment containing the *lac* promoter is favored by 5 kcal/mole over binding to fragments that do not contain a binding site (*24*). However, in comparison to the *lac* site at −60, binding of CRP to the *mal*T site is less stable by roughly 0.6 kcal/mole, to the *gal* site by 1.4 kcal/mole, and to the CRP site in the *lac* operator by 2.5 kcal/mole (*21*). In these cases only one CRP dimer is bound per site (*18, 19, 21*), and a single cyclic AMP molecule bound to the dimer suffices for the stable binding of the complex (*25*). This hierarchy of CRP binding affinities to different DNA templates may explain the observation that different cyclic AMP–sensitive genes are activated by different concentrations of cyclic AMP in vivo (*26, 27*).

5) The distance between the transcription start point and the CRP binding site is different at different promoters (see Table 1). Furthermore, the 5′TGTGA3′ sequence, which is always located toward one end of the zone protected by deoxyribonuclease, is found either at the end distal or proximal to the transcription start (see Fig. 1). The motif can, therefore, be found on either the "transcribed" or "nontranscribed" strand. Hence, when the sequence recognized by CRP is asymmetric, the consensus sequence can adopt either of two orientations with respect to the direction of transcription. These differences in location and orientation of the CRP sites indicate that the distance between the CRP site and the start of transcription is not critical, and argues against a fixed interaction between cyclic AMP–CRP and the RNA polymerase molecule in open complexes, analogous to the interactions of the sigma subunit with the rest of the enzyme.

6) At some promoters, like those for the *ara*BAD, *mal*K-*lam*B, and *mal*E, F, and G operons, an additional protein, the AraC protein or the MalT protein is also required for activation of transcription (*28*). These activator proteins also bind to the promoter, but it is not known whether there are direct protein interactions between the two activating proteins.

Table 1. Distance (in number of bases) between center of CRP binding sites and start of transcription.

*mal*E	−106/107	*ilv*B	−64/65
*mal*K	−99/100	pBR-P4	−42/43
*ara*BAD	−93/94	*gal*	−41/42
*mal*T	−70/71	*deo* site 1	−40/41
lac site 1	−61/62	*omp*A	−36/37
*tna*A	−61/62		

7) In some cases (for example, *lac*, *ara*, *cat*) two CRP binding sites have been found (*5, 6*). The secondary sites bind CRP less tightly than the primary sites. Although there is no evidence that they are essential for transcription activation, they may participate in the search of CRP for the main site.

8) Examination of the sequences around the transcription start points at CRP regulated promoters reveals a particularly poor correspondence with the consensus sequences that have been established for *Escherichia coli* promoters. In most cases, the −35 sequence "TTGACA" is barely apparent, the Pribnow box sequence ("TATAAT") is often very imperfect, and the spacing between these two elements is more than the optimal 17 bp. This is perhaps not surprising because a good fit with these three structural determinants would result in a very efficient promoter, which presumably would be very difficult to regulate by CRP. Moreover, the CRP-dependent promoter sequences show no apparent systematic deviation from the consensus.

9) Although the major role of cyclic AMP–CRP is undoubtedly to activate transcription, CRP binding can in some systems repress transcription. At the *gal* promoter, CRP binding is responsible not only for the activation of one promoter (P1) but also for the inhibition of a second (P2) (*29*). Because CRP binds to the −35 region of the P2 promoter, it is likely that CRP blocks access of RNA polymerase to P2 (*4*). CRP also acts as a repressor of transcription of its own promoter in vitro. Similarly, in vitro, CRP inhibits transcription of the gene for the major outer membrane protein, OmpA, again by binding near the −35 region of the promoter (*8*).

Structure of CRP

CRP is a dimer composed of two identical subunits, each containing 210 amino acids. Its complete amino acid sequence has been deduced from the nucleotide sequence of the cloned CRP gene (*30*). Equilibrium dialysis studies indicate that two molecules of cyclic AMP can bind per CRP dimer (*31*).

Like several other DNA-binding regulatory proteins, CRP is composed of a DNA-binding domain and of another domain that acts to regulate binding to DNA. For CRP the existence of two

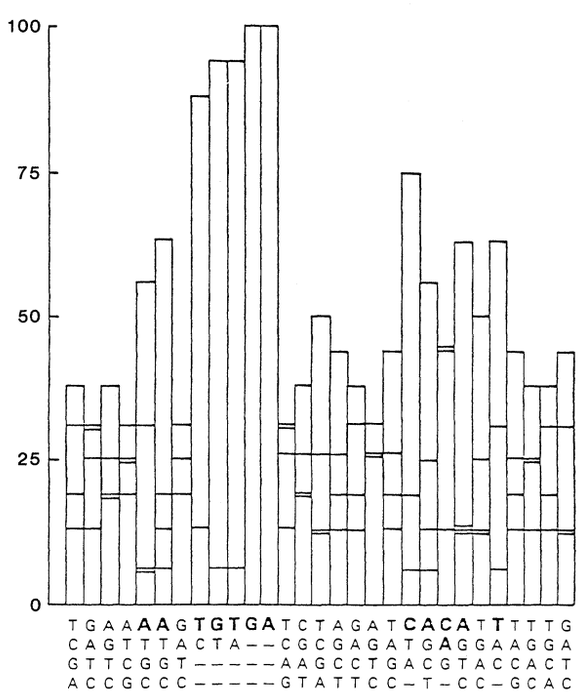

Fig. 2. Histogram of sequences in CRP binding sites. The CRP sites are aligned on the TGTGA motif as in Fig. 1. For each position the base marked on the first line is represented by the highest bar and is expressed in percent occurrence; the base on the second line is represented by the second highest bar, and so on.

separate domains was first suggested by a series of proteolytic digestion experiments that identified an amino-proximal fragment that retains cyclic AMP–binding activity but has lost DNA–binding activity and a carboxyl-terminal fragment that retains DNA binding but has lost cyclic AMP binding (*32*).

This two-domain structure was clearly shown by the 2.9 Å resolution crystal structure of CRP complexed with cyclic AMP (see Fig. 3). A larger amino-terminal domain extending from residue 1 to 135 is separated by a cleft from a smaller carboxyl-terminal domain that extends from residue 136 to 210 (*33*). The amino-terminal domain of each subunit contains one cyclic AMP molecule buried in the interior of the protein. Residues from both subunits appear to be involved in the binding of each cyclic AMP. The 6-amino group of the adenine ring of cyclic AMP interacts with Thr^{127} (threonine at residue 127) on one subunit and Ser^{128} on the other (*33, 34*).

A segment of the amino-terminal domain of CRP (residues 30 to 89) exhibits significant sequence homologies with two segments of the regulatory subunit, R-II of the eukaryotic cyclic AMP–dependent protein kinase enzyme. Since each R-II subunit contains two cyclic AMP–binding sites, it is possible that R-II contains two structural domains similar to the β-roll structure in CRP, in which cyclic AMP is buried when it interacts with the protein. Hence, it has been proposed that a specifically conserved cyclic AMP–binding domain may be used in both proteins to respond to the same physiological stimulus (*35*). In bacteria, this domain is linked to a DNA binding domain in a single polypeptide chain and influences the DNA binding domain by causing allosteric conformational changes; whereas in eukaryotes it interacts through subunit-subunit interactions with the catalytic subunit of the protein kinase.

The smaller carboxyl-terminal domain of each CRP subunit consists of three α-helices connected by short β-sheet structures. In each subunit, one of these α-helices, F, clearly protrudes from the surface of the CRP dimer. These two α-helices are thought to provide the major interaction sites with the DNA targets since their axes run approximately parallel to each other at a distance of 34 Å (*33*). Recently CRP mutants were isolated that suppress two symmetrically located *lac* CRP-binding site mutations. The CRP mutations cause a change in one amino acid in the N-terminal part of α-helix F, implying that this segment provides crucial interactions with the DNA template (*35a*).

Comparison of CRP with Other DNA Binding Poteins

The understanding of the structure of CRP was enhanced by comparing it with the x-ray structures of the Cro and cI repressors of bacteriophage lambda (*36, 37*). This comparison reveals a striking structural homology in the relative position and orientation of two consecutive α-helices, helix E and helix F, in the carboxyl-terminal domain of CRP, and helix II and III in the amino-terminal part of Cro. Moreover, the path of the 24 α-carbon atoms in the structural unit that contains these two helices in CRP can be superimposed on the path of the α-carbon atoms of the two homologous helices in Cro (*38*). A similar structural homology is also seen between the same two helices in Cro and two similar helices in the lambda cI repressor (*37*). Thus a conserved helix-turn-helix motif is found in each of the three proteins. Furthermore, in a number of other DNA binding proteins sequences appear which show a clear homology with those that form the helix-turn-helix motif in CRP, Cro and cI (*39, 40*). These proteins, for which x-ray crystallographic data are not yet available, include the cII proteins of phages lambda and 434, the repressor and Cro proteins of phages P22 and 434, and the *lac* and *gal* repressors of *E. coli*. Some *lac* repressor mutations that are defective in operator binding map in the sequence that is homologous to the conserved helix-turn-helix motif (*41*).

All these DNA binding proteins contain, therefore, a similar helix-turn-helix domain that is probably essential for their interactions with DNA. During the assembly of these genes, sequences for additional domains were probably added to a basic DNA recognition unit. These additional domains provide the determinants required for ligand binding, subunit-subunit interactions, and allosteric modifications of the protein. Although the basic design of the DNA interaction unit appears to be conserved, most of these different proteins recognize a distinct site on the DNA. The specificity in DNA recognition must be provided by the side chains in the motif and possibly by other adjacent residues for both repressors and activators.

A striking difference between CRP and the other proteins is that in CRP it is the carboxyl-terminal part of the molecule that contains the helix-turn-helix DNA binding motif, whereas in the other DNA binding proteins discussed above, this motif is located in the amino-proximal part of the molecule. Recently, however, a study of the sequence of the *E. coli fnr* gene, which is essential for anaerobic respiratory metabolism, suggested that in this case it is also the carboxyl-terminal end that interacts with DNA (*42*). Indeed, a sequence homologous to the helix-turn-helix motif is found at about the same place in the Fnr protein as in CRP. Additional areas of homology are found between the amino-proximal domain of CRP and the Fnr protein although the Fnr protein presumably does not bind cyclic AMP. Interestingly, the Fnr protein fulfills a somewhat similar function as CRP: it is a pleiotropic activator for a series of genes which are turned on when aerobic metabolism becomes limiting. The sequence analysis suggests that the *fnr* gene may have derived by duplication either from the CRP gene itself or from a common ancestor.

Possible Interactions of CRP with DNA Deduced from X-ray Structure

For both the lambda Cro protein (*36, 43*) and the lambda cI protein (*37*) attractive models for interactions with DNA were suggested from the x-ray data. The models suggest that two symmetrically related α-helices in each of these dimeric proteins interact with two symmetrical DNA sequences in two successive major grooves of right-handed DNA. In each of these two proteins, as in CRP, this interacting α-helix in one subunit is separated from the symmetrical helix in the other subunit by a distance of 34 Å. The models account for both the biochemical protection data and the location of mutations in the target DNA. For the Cro dimer, the two α-helices fit neatly in two successive major grooves of right-handed DNA (*36, 43*); for the lambda repressor, the relative positions of these two α-helices in the major grooves are somewhat different (*37*). For CRP, the x-ray data were initially interpreted to rule out a model of interactions with right-handed DNA analogous to those proposed for Cro and cI (*34*). Indeed, if the long axis of the DNA binding domain in the dimer is aligned on the long axis of the DNA, the two F α-helices appear to lie across rather than parallel to the major grooves of right-handed DNA. Because the relative orientation of these two F helices is complementary to two successive major grooves of left-handed B DNA, McKay *et al.* proposed that CRP binds to left-handed B DNA (*33*). This is, however, very unlikely, because if the DNA at the CRP target site was left-handed, the binding of cyclic AMP–CRP to a closed circular DNA would greatly change its linking number in experiments where the DNA is nicked, then resealed. In fact,

only very small changes in the linking number were found when the cyclic AMP–CRP complex binds to a closed supercoiled circular DNA containing either the *lac* or the *gal* CRP binding site (*44*).

An alternative model where CRP interacts with two successive major grooves of right-handed B DNA was, therefore, considered (*37, 40, 45*). This model proposes that it is the NH_2-terminal part of the F α-helices which enters two successive major grooves in a way that is similar to the interaction of helix "3" of the lambda repressor. For the lambda repressor the orientation of helix "3" is thought not to be exactly parallel to the path of the major groove (*37*). The two symmetrical helices "3" of the cI dimer cannot be fit completely in two successive major grooves, mainly because the carboxyl-terminal part of these helices do not protrude enough from the surface of the protein. In the case of CRP a similar arrangement would allow interactions for about 8 to 9 bp in a region of 14 bp, shorter than the CRP binding site. However, a bend in the DNA would provide an opportunity for additional contacts with the DNA backbone (*45*). Recent experiments strongly suggest that binding of CRP to DNA causes the DNA to bend (*46*). A further alternative hypothesis is that the active form of CRP is only obtained after the cyclic AMP–CRP complex interacts with DNA and that this active form has a different configuration from that which was crystallized. X-ray analysis of cocrystals of CRP with DNA should obviously tell whether this is the case.

Multiple Conformational States of CRP

Several types of experiments indicate that the binding of cyclic AMP to CRP causes conformational changes in the protein. For example, the sensitivity of CRP to cleavage by proteolytic enzymes is much higher in the presence of cyclic AMP (*32*). Another example is the induction by dithiobis-(2-nitrobenzoic acid) (DTNB) of an intersubunit disulfide cross-link in CRP that occurs only in the presence of cyclic AMP (*32*). Experiments with a mutant CRP, which is able to function in the absence of cyclic AMP, suggest that these changes are important for activation of transcription. This mutant CRP, indeed, displays a sensitivity to proteolytic enzymes in the absence of cyclic AMP which resembles the sensitivity of wild-type CRP in the presence of cyclic AMP (*47*). This suggests that the conformation of the mutant protein in the absence of cyclic AMP is similar to the one that is normally induced by cyclic AMP in wild-type CRP.

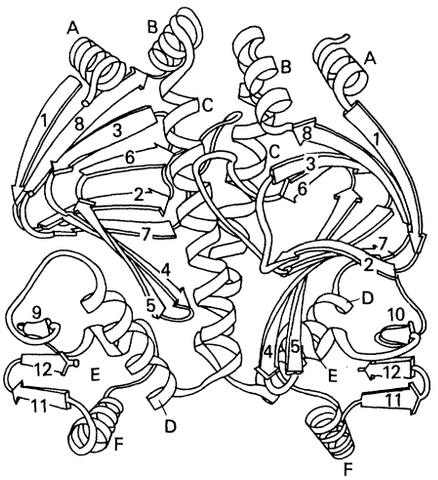

Fig. 3. Drawing of the CRP dimer. The amino-proximal domain consists of α-helix A, β-sheets 1 to 8, and α-helices B and C. The DNA binding carboxyl-terminal domain consists of α-helices D, E, and F, and the residues connecting these helices. The two F helices, which clearly protrude from the dimer, are thought to provide many of the interactions with DNA. All the interactions between the two subunits are provided by the large amino-terminal domain and the majority of these are provided by the two long C helices that lie together in the center of the dimer. The two subunits are not exactly related by a perfect dyad axis of symmetry. [From (*40*); courtesy of Cold Spring Harbor Laboratory Press].

A minimal model to explain these conformational changes would suggest that two conformations exist, an "inactive" conformation and an "active" conformation which is more sensitive to proteolysis and disulfide cross-linking, which binds cyclic AMP, and which has an increased affinity for DNA. Indeed, the addition of cyclic AMP does strongly increase the overall affinity of CRP for double-stranded DNA (*48, 49*). Conversely, binding to DNA increases the affinity of cyclic AMP for CRP (*31*). However, this two-state model fails to account for all the data. Indeed, some of its predictions are clearly not fulfilled. (i) All the changes in the protein structures due to one activator (cyclic AMP) should also be triggered by the other (DNA). However, DNA increases the fluorescence quantum yield of a tryptophan residue which is unaffected by cyclic AMP binding (*24*). Further, DNA causes the protein to aggregate beyond the dimeric state (*50–52*). These cooperative effects are clearly dependent on the binding of CRP to DNA because in its absence CRP does not show any tendency to aggregate beyond its dimeric structure. (ii) Only two types of cyclic AMP analogs should be found, those that can trigger the overall activating process and those like cyclic guanosine monophosphate (GMP) unable to do it. In fact, another class of analogs with monosubstitutions at the N-6 or C-2 positions in the purine bind to CRP as well as cyclic AMP and produce the same effects on proteolysis and disulfide cross-linking. However, these analogs fail to stimulate binding to poly(dA-T) (polydeoxyadenylate-deoxythymidylate) or to promoter DNA and to activate transcription in vitro (*54*). Cyclic AMP must therefore cause an additional conformational modification in CRP that this other class of analogs is unable to produce.

Changes in DNA Structure Caused by CRP

By interacting with DNA, CRP alters the structure of DNA. Circular dichroism studies are consistent with the notion that in both the presence and absence of cyclic AMP, the DNA becomes more compact (*50, 51*). In experiments performed in the absence of cyclic AMP, molecules of pBR322 DNA were condensed fourfold when covered by CRP, resulting in the formation of long cylindrical rods of approximately 110 Å cross-section with regular periodical striations. In these structures the DNA is probably arranged as a tightly wound solenoidal supercoil (*52, 53*).

Structural changes in the DNA are also found when cyclic AMP–CRP interacts with specific DNA sites. Indeed, a 1:1 complex between a *lac* promoter DNA fragment and cyclic AMP–CRP was found to display an anomalous electrophoretic mobility, probably caused by structural changes in the DNA that are induced by CRP (*20, 29*). It is clear that these changes are not due to a transition from right-handed to left-handed DNA (*44*), nor are they due to unwinding of the DNA because binding of cyclic AMP–CRP to its recognition site at the *lac* promoter produces, in fact, a stabilization of the DNA duplex (*55*). However, the significance of the changes in DNA structure for the activation of transcription by cyclic AMP–CRP remains unexplained.

Activation of Transcription by Cyclic AMP–CRP

Much of our understanding of transcription initiation in bacteria is based on a comparison of DNA sequences of wild-type and mutant promoters and on various assays designed to measure the functional strength of these promoters.

Examination of different promoter se-

quences shows two regions of homology, one centered at about −10, the other in the −35 region (*56*). The distance between the −35 consensus and the −10 consensus is optimally 17 bp (*57*), but promoters with spacing of as little as 15 or as many as 20 bp retain a partial function. Strong promoters show a higher degree of homology with these consensus sequences than do weaker promoters. Although the vast majority of promoter mutants map in one or the other of the two consensus sequences, some map outside the −10 or −35 homologies, an indication that structural features other than the −10 and −35 consensus may also play a role in determining promoter strength.

Interactions between RNA polymerase and promoters can generally be described by a simple two-step model (*58*). In this scheme, the enzyme binds to the promoter to form an inactive or "closed" complex. This binding step is reversible and is characterized by an association constant, K_B. The closed complex isomerizes to give rise to an active or open complex. This isomerization which includes a localized unwinding of the DNA over a distance of approximately 12 bp near the transcription start is generally irreversible, and the corresponding rate constant, k_f, is slow (*59*).

McClure has developed a method, called the abortive initiation assay, which allows quantitation of both K_B and k_f (*59*). This method is based on the ability of RNA polymerase, when present in an open complex, to synthesize, at a steady-state rate, a short oligonucleotide corresponding to the 5′ end of a promoter specific messenger RNA (mRNA). Elongation is prevented by the omission of one or more nucleoside triphosphates. If the reaction is started by the addition of RNA polymerase, a lag occurs before the establishment of the rate of the steady state. A kinetic analysis of this lag at different RNA polymerase concentrations gives information about both the association constant of the binding step and the rate constant of the isomerization step (K_B and k_f). The final steady-state rate of oligonucleotide synthesis is a measure of the number of promoters at which an RNA polymerase molecule is correctly placed (*59*).

The usefulness of the abortive initiation assay has been substantiated by a series of results. First, strong promoters have high values of both K_B and k_f, whereas weak promoters have low values for both constants. Second, the phenotype of both up and down promoter mutants can be explained from their behavior in vitro because the mutations affect one of the two constants or both. In addition, a comparative study of a number of wild-type and mutant promoters has identified these elements in the promoter sequence that are important in determining the values of K_B and k_f. A high value of K_B is generally associated with a good consensus sequence in the −35 region and, conversely, mutants in the −35 sequence affect K_B. The value of k_f depends mainly, but not exclusively, on the degree of homology of the −10 sequence with the consensus Pribnow sequence, and on the distance between the −35 and the −10 sequences (*57, 60*).

CRP Activation of the *lac* Promoter

The abortive initiation assay has been used to examine the role of CRP with the *lac* promoter as template (*61*). The wild-type *lac* promoter is a very poor promoter in the absence of CRP and cyclic AMP, and the values of both k_f and K_B are low. Even after long incubation times only 3 percent of the wild-type *lac* promoter molecule actively transcribes the 5′ *lac* mRNA oligonucleotide, suggesting that other sites in the *lac* control region may be competing for the polymerase. Interestingly, a second *lac* promoter, P2, which promotes transcription 22-bp upstream from the conventional *lac* transcription start point, has recently been identified from in vitro transcription experiments (*61, 62*). Thus the simplest explanation for the low fractional occupancy of the *lac* promoter (P1) in the absence of cyclic AMP–CRP is that the *lac* P2 promoter and maybe other cryptic promoter sites act as competitors for RNA polymerase binding and divert the RNA polymerase molecules away from the conventional *lac* promoter (P1). Because the P2 and P1 sequences overlap each other, RNA polymerase bound to P2 will block access to P1.

Addition of cyclic AMP and CRP has two effects on the *lac* promoter. First, it enhances the rate of open complex formation. This occurs by increasing the value of K_B without affecting k_f. Second, the presence of cyclic AMP–CRP also increases the fractional occupancy of the *lac* P1 promoter by RNA polymerase. This increased occupancy is due to inhibition by cyclic AMP–CRP of RNA polymerase binding to P2 and perhaps to other secondary sites. Thus, in the presence of the cyclic AMP–CRP complex, the P2 promoter no longer competes for RNA polymerase binding to P1, and RNA polymerase molecules are no longer diverted from forming an open complex at the *lac* promoter (*61*).

The inhibition by cyclic AMP–CRP of RNA polymerase binding to the *lac* P2 site and to other sites does not explain the observed increase in K_B (*61*). Indeed, if inhibition of RNA polymerase binding at P2 was solely responsible for the effect of the complex on *lac* transcription, the overall binding constant of RNA polymerase, K_B, would be expected to decrease because, by merely blocking some binding sites, cyclic AMP–CRP should decrease the total number of reversible interactions between RNA polymerase and the *lac* promoter. Therefore, the cyclic AMP–CRP complex has both a direct (stimulation of K_B) and an indirect effect (inhibition of RNA polymerase binding to accessory promoters).

CRP Activation of the *gal* Promoter

Dual control by cyclic AMP–CRP of two overlapping promoters was first observed in the *gal* operon (*29*). One promoter (P1) is dependent on the presence of cyclic AMP–CRP whereas the second promoter (P2) is active in the absence of cyclic AMP–CRP and is inhibited by these factors. Increasing the concentration of CRP or cyclic AMP causes an increase in P1 and a proportional decrease in P2 activity. In cells harboring a point mutation in the *gal* CRP binding site, cyclic AMP–CRP is unable either to inhibit P2 or to activate P1; thus, the two effects must be due to the binding of CRP at the same site (*15*). The two promoters overlap each other and are mutually exclusive. The striking analogy of the two *gal* promoters with the two *lac* promoters suggests that a similar mechanism of competing and mutually exclusive promoters is a necessary part in the control of activation of transcription initiation at both the *lac* and *gal* P1 promoters. An analysis of the kinetics of open complex formation at the *gal* P1 and P2 promoters again shows that the same cyclic AMP–CRP complex not only excludes RNA polymerase from P2 but also acts positively on the kinetic parameters of P1 (*63*).

Mechanism of Activity and Its Implications

Models to explain how CRP activates transcription initiation must take into account the following points.

1) In *lac*, cyclic AMP–CRP does not affect the isomerization step in the RNA polymerase-promoter interaction, but

stimulates the formation of closed complexes between RNA polymerase and promoter DNA that have a high probability of forming open complexes and initiate transcription at the correct site. By analogy, cyclic AMP–CRP may have a similar function in the other cyclic AMP–dependent promoters; activation of transcription is achieved by increasing the affinity of RNA polymerase for the promoter and by blocking competing and overlapping promoter sites.

2) The distance between the CRP binding site and the start of transcription is not the same at different cyclic AMP–dependent promoters; and, when the CRP binding site is asymmetric, the orientation relative to the direction of transcription also varies. It is possible that the distance between the site where RNA polymerse makes its initial interactions with the promoter in a closed complex and the transcription start point is not crucial.

3) The cyclic AMP–CRP complex causes changes in the DNA structure. These changes are not transitions from right-handed to left-handed DNA. Cyclic AMP–CRP also does not unwind the DNA duplex and, therefore, does not stimulate transcription by unwinding neighboring regions of the promoter. The step at which CRP acts at the *lac* promoter, that is, the formation of a correct closed complex, would not require such changes in the DNA. However, we do not know whether the observed changes in DNA structure are needed for the activation process.

4) Numerous indirect arguments favor the view that protein-protein interactions between cyclic AMP–CRP complex and RNA polymerase could mediate the increase in K_B. Indeed, in the presence of cyclic AMP, RNA polymerase and CRP have been shown to cosediment (*64*). Furthermore, the increased affinity of the cyclic AMP–CRP complex for its binding site at the *lac* and *gal* promoter in the presence of RNA polymerase (*63*) is also consistent with this view. There is evidence that CRP remains bound to its site on *gal* DNA at −35 after formation of the open complex, and that CRP and RNA polymerase are very close to each other in the ternary complex (*65*). The location of the *gal* CRP site around −35 and the absence of a −35 sequence for *gal* P1 (the cyclic AMP–CRP dependent promoter) have prompted the hypothesis that the CRP protein itself provides some of the interactions normally provided by the −35 segments in other promoters (*15*). Since the other cyclic AMP–sensitive promoters also show a generally poor homology with the −35 consensus sequence CRP may somehow also be able to compensate for this deficiency and orient the RNA polymerase molecule in its search for a correct Pribnow box. It is, therefore, possible that the CRP dimer contains, in addition to its cyclic AMP and DNA binding sites, a site for interactions with RNA polymerase.

Models for the mechanism of activation of transcription are inherently much more difficult to formulate (but more interesting to study) than models for repression of transcription. Indeed, if a repressor has a sufficient affinity for a given DNA segment at a promoter, its binding to this fragment will sterically prevent the RNA polymerase from making the correct interactions with the promoter. To be able to undergo positive regulation over a wide range of activities a promoter needs, first, to have a low level of activity in the absence of the activator. This can be accomplished, for example, if the sequence determinants that are important for RNA polymerase–promoter interaction deviate from the consensus sequence or if additional RNA polymerase binding sites are present in the promoter which can divert the enzyme from initiating at the correct start site. Activation could occur by stimulating either the binding of RNA polymerase or its isomerization rate, or by blocking competing RNA polymerase binding sites. The lambda cI protein, which is an activator for the synthesis of its own RNA, activates transcription of the P_{RM} promoter of lambda DNA by increasing the isomerization rate without affecting the binding constant, K_B (*66*). In contrast, cyclic AMP–CRP activates *lac* transcription both by increasing the binding constant of RNA polymerase and by excluding competing RNA polymerase binding sites. How the increased binding is accomplished by CRP is still unknown, largely because of our ignorance about the biochemical properties of closed complexes. It seems clear, however, that the isolation and biochemical characterization of different CRP mutants should help resolve this problem.

References and Notes

1. W. Epstein, L. B. Rothman-Denes, J. Hesse, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2300 (1975).
2. R. Perlman and I. Pastan, *Biochem. Biophys. Res. Commun.* **37**, 151 (1969); D. Schwartz and J. Beckwith, in *The Lactose Operon*, J. Beckwith and D. Zipser, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1970), p. 417; E. Brickman, L. Soll, J. Beckwith, *J. Bacteriol.* **116**, 582 (1973).
3. J. Majors, thesis, Harvard University (1977); R. Simpson, *Nucleic Acids Res.* **8**, 759 (1980).
4. T. Taniguchi, M. O'Neill, B. de Crombrugghe, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5090 (1979).
5. S. Ogden, D. Haggerty, C. Stoner, D. Kolodrubetz, R. Schleif, *ibid.* **77**, 3346 (1980); N. Lee *et al.*, *ibid.* **78**, 752 (1981).
6. S. Le Grice and H. Matzura, *J. Mol. Biol.* **150**, 185 (1981); A. Schmitz, *Nucleic Acids Res.* **9**, 277 (1981); S. Le Grice, H. Matzura, R. Marcoli, S. Iida, T. Bickle, *J. Bacteriol.* **150**, 312 (1982); P. Valentin-Hansen, *EMBO J.* **1**, 1049 (1982); P. Friden, T. Newman, M. Freundlich, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6156 (1982).
7. C. Queen and M. Rosenberg, *Nucleic Acids Res.* **9**, 3365 (1981).
8. R. Movva, P. Green, K. Nakamura, M. Inouye, *FEBS Lett.* **128**, 186 (1981); H. Aiba, *Cell* **32**, 141 (1983).
9. B. de Crombrugghe, B. Chen, W. Anderson, P. Nissley, M. Gottesman, I. Pastan, *Nature (London) New Biol.* **231**, 139 (1971); L. Eron and R. Block, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1828 (1971); S. P. Nissley, W. Anderson, M. Gallo, I. Pastan, R. Perlman, *J. Biol. Chem.* **246**, 4671 (1971); B. de Crombrugghe, I. Pastan, W. Shaw, J. Rosner, *Nature (London) New Biol.* **241**, 237 (1973); G. Wilcox, P. Meuris, P. Bass, E. Englesberg, *J. Biol. Chem.* **249**, 2946 (1974); S. Le Grice and H. Matzura, *FEBS Lett.* **110**, 11 (1980).
10. M. C. Deeley and C. Yanofsky, *J. Bacteriol.* **151**, 942 (1982).
11. P. Valentin-Hansen, H. Aiba, D. Schumperli, *EMBO J.* **1**, 317 (1982).
12. J. Hopkins, *J. Mol. Biol.* **87**, 715 (1974).
13. J. Majors, *Nature (London)* **256**, 672 (1975).
14. R. Dickson, J. Abelson, R. Johnson, W. Reznikoff, W. Barnes, *J. Mol. Biol.* **111**, 65 (1977).
15. S. Busby, H. Aiba, B. de Crombrugghe, *ibid.* **154**, 211 (1982); S. Busby and M. Dreyfus, *Gene* **21**, 121 (1983).
16. C. G. Miyada, X. Soberon, K. Itakura, G. Wilcox, *Gene* **17**, 167 (1982).
17. M. C. O'Neill, K. Amass, B. de Crombrugghe, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2213 (1981); R. Ebright, in *Molecular Structure and Biological Activity*, J. Griffen and W. Duax, Eds. (Elsevier, New York, 1982), pp. 91–100.
18. A. Kolb, S. Busby, M. Herbert, D. Kotlarz, H. Buc, *EMBO J.* **2**, 217 (1983); C. Chapon and A. Kolb, *J. Bacteriol.* **156**, 1135 (1983).
19. M. Garner and A. Revzin, *Biochemistry* **21**, 6032 (1982).
20. M. Fried and D. Crothers, *Nucleic Acids Res.* **11**, 141 (1983); D. Crothers and M. Fried, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 263 (1983).
21. A. Kolb, A. Spassky, C. Chapon, B. Blazy, H. Buc, *Nucleic Acids Res.* **11**, 7833 (1983).
22. We cannot formally exclude the existence of two alternative nonsymmetrical arrangements of the CRP dimer at this site if these two arrangements are equally frequent.
23. S. Busby, D. Kotlarz, H. Buc, *J. Mol. Biol.* **167**, 259 (1983).
24. M. Takahashi, B. Blazy, A. Baudras, W. Hillen, *ibid.*, p. 895.
25. S. Shanblatt and A. Revzin, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1544 (1983).
26. J. Lis and R. Schleif, *J. Mol. Biol.* **79**, 149 (1973).
27. M. Piovant and C. Lazdunski, *Biochemistry* **14**, 1821 (1975).
28. N. Lee, in *The Operon*, J. Miller and W. Reznikoff, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1978), pp. 389–409; M. Schwartz, *Ann. Inst. Pasteur (Paris)* **113**, 685 (1967).
29. R. Musso, R. Di Lauro, S. Adhya, B. de Crombrugghe, *Cell* **12**, 847 (1977).
30. W. B. Anderson, A. B. Schneider, M. Emmer, R. L. Perlman, I. Pastan, *J. Biol. Chem.* **246**, 5929 (1971); H. Aiba, S. Fujimoto, N. Ozaki, *Nucleic Acids Res.* **10**, 1345 (1982); P. Cossart and B. Gicquel-Sanzey, *ibid.*, p. 1363.
31. T. Takahashi, B. Blazy, A. Baudras, *Biochemistry* **19**, 5124 (1980).
32. J. Krakow and I. Pastan, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2529 (1973); E. Eilen and J. Krakow, *J. Mol. Biol.* **114**, 47 (1977): E. Eilen, C. Pampeno, J. S. Krakow, *Biochemistry* **17**, 2469 (1978); C. Pampeno and J. Krakow, *ibid.* **18**, 1519 (1979); H. Aiba and J. Krakow, *ibid.* **20**, 4774 (1981).
33. D. McKay and T. Steitz, *Nature (London)* **290**, 744 (1981); D. McKay, I. Weber, T. Steitz, *J. Biol. Chem.* **257**, 9518 (1982).
34. The x-ray data predict that it is the anti-conformation of cyclic AMP that interacts with CRP. However, a nuclear magnetic resonance study of CRP suggests that it is the *syn* conformation of the cyclic nucleotide that interacts with CRP in solution [A. M. Gronenborn and G. M. Clore, *Biochemistry* **21**, 4040 (1982)]. This paradox has not yet been resolved.
35. I. Weber, K. Takio, K. Titani, T. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7679 (1982).

35a. R. H. Ebright, P. Cossart, B. Gicquel-Sanzey, J. Beckwith, personal communication.

36. W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. Matthews, *Nature (London)* **290**, 754 (1981).
37. C. Pabo and M. Lewis, *ibid.* **298**, 443 (1982).
38. T. Steitz, D. Ohlendorf, D. McKay, W. Anderson, J. B. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3097 (1982).
39. W. F. Anderson, Y. Takeda, D. H. Ohlendorf, B. W. Matthews, *J. Mol. Biol.* **159**, 745 (1982); B. Gicquel-Sanzey and P. Cossart, *EMBO J.* **1**, 591 (1982); B. M. Matthews, D. H. Ohlendorf, W. F. Anderson, Y. Takeda, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1428 (1982); R. Sauer, R. Yocum, R. Doolittle, M. Lewis, C. Pabo, *Nature (London)* **298**, 447 (1982); I. Weber, D. McKay, T. Steitz, *Nucleic Acids Res.* **10**, 5085 (1982).
40. T. Steitz, I. T. Weber, J. B. Matthews, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 419 (1983).
41. K. Adler, K. Beyreuther, E. Fanning, N. Geizler, B. Gronenborn, A. Klemm, B. Muller-Hill, M. Pfahl, A. Schmitz, *Nature (London)* **237**, 322 (1972).
42. D. J. Shaw, D. W. Rice, J. R. Guest, *J. Mol. Biol.* **166**, 241 (1983).
43. D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, B. W. Matthews, *Nature (London)* **298**, 718 (1982).
44. A. Kolb and H. Buc, *Nucleic Acids Res.* **10**, 473 (1982).
45. T. Steitz and I. Weber, in *Structural Biology*, A. McPherson and F. Jurnak, Eds. (Wiley, New York, in press).
46. H. Buc, in preparation.
47. J. G. Harman and W. J. Dobrogosz, *J. Bacteriol.* **153**, 191 (1983).
48. M. Takahashi, B. Blazy, A. Baudras, *Nucleic Acids Res.* **7**, 1699 (1979).
49. S. P. Nissley, W. Anderson, M. Gallo, I. Pastan, R. Perlman, *J. Biol. Chem.* **247**, 4264 (1971); A. D. Riggs, G. Reiness, G. Zubay, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1222 (1971).
50. S. Saxe and A. Revzin, *Biochemistry* **18**, 255 (1979).
51. S. R. Martin, A. M. Gronenborn, G. M. Clore, *FEBS Lett.* **159**, 102 (1983); B. Blazy, personal communication.
52. J.-J. Chang, J. Dubochet, A. Baudras, B. Blazy, M. Takahashi, *J. Mol. Biol.* **150**, 435 (1981).
53. F. R. Salemme, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5263 (1982).
54. R. Ebright, J. Krakorw, S. Le Grice, J. Miller, personal communication.
55. B. Unger, G. M. Clore, A. M. Gronenborn, W. Hillen, *EMBO J.* **2**, 289 (1983).
56. M. Rosenberg and D. Court, *Annu. Rev. Genet.* **13**, 319 (1979); U. Siebenlist, R. B. Simpson, W. Gilbert, *Cell* **20**, 269 (1980); D. K. Hawley and W. R. McClure, *Nucleic Acids Res.* **11**, 2237 (1983).
57. J. Stefano and J. Gralla, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1069 (1982).
58. G. Walter, W. Zillig, P. Palm, E. Fuchs, *Eur. J. Biochem.* **3**, 194 (1967); M. Chamberlin, *Annu. Rev. Biochem.* **43**, 721 (1974).
59. W. R. McClure, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5634 (1980).
60. ——— and D. K. Hawley, in *Mobility and Recognition in Cell Biology*, H. Sund and C. Veeger, Eds. (de Gruyter, Berlin, 1982), p. 317.
61. D. K. Hawley, T. P. Malan, M. Mulligan, W. McClure, in *Promoters: Structure and Function*, R. Rodriguez and M. Chamberlin, Eds. (Praeger, New York, 1982), pp. 54–68; W. R. McClure, D. K. Hawley, T. P. Malan, in *ibid.*, pp. 111–140; T. P. Malan, A. Kolb, H. Buc, W. McClure, in preparation.
62. W. S. Reznikoff, L. E. Maquat, L. M. Munson, R. C. Johnson, W. Mandecki, in *Promoters: Structure and Function*, R. Rodriguez and M. Chamberlin, Eds. (Praeger, New York, 1982), pp. 80–95.
63. M. Herbert, A. Spassky, A. Kolb, S. Busby, H. Buc, unpublished observations.
64. B. Blazy, M. Takahashi, A. Baudras, *Mol. Biol. Rep.* **6**, 39 (1980).
65. T. Taniguchi and B. de Crombrugghe, *Nucleic Acids Res.* **11**, 5165 (1983).
66. D. Hawley and W. R. McClure, *J. Mol. Biol.* **157**, 493 (1982).

RESEARCH ARTICLE

Light-Regulated Expression of a Pea Ribulose-1,5-Bisphosphate Carboxylase Small Subunit Gene in Transformed Plant Cells

Richard Broglie, Gloria Coruzzi, Robert T. Fraley
Stephen G. Rogers, Robert B. Horsch, Jeanne G. Niedermeyer
Cynthia L. Fink, Jeffery S. Flick, Nam-Hai Chua

The enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco) (E. C. 4.1.1.39) catalyzes the fixation of carbon dioxide in photosynthetic organisms. In higher plant chloroplasts the holoenzyme is composed of eight copies each of two nonidentical subunits; a large subunit (rbcL) encoded by chloroplast DNA (*1, 2*) and a small subunit (rbcS) encoded by nuclear DNA (*3*). Recent chloroplast transport studies in vitro have demonstrated that the rbcS polypeptide is synthesized as a larger precursor that is imported into chloroplasts by an energy-dependent (*4*), posttranslational process (*3*). The genes coding for rbcS exhibit a diverse array of regulatory properties. In peas, the rbcS polypeptide is a major product of cytoplasmic protein synthesis in leaves, but is either absent or present in reduced amounts in other plant parts (*5*). In plants that utilize the C_4 photosynthetic pathway, both rbcS and rbcL subunits are present in bundle sheath cells but are absent from mesophyll cells (*6*). In tissues containing Rubisco, the expression of the nuclear and chloroplast genes encoding the polypeptide subunits are controlled by light (*5, 7*), and this effect is mediated by phytochrome (*8*).

Complementary DNA (cDNA) clones encoding rbcS have been isolated from peas (*9, 10*) and wheat (*11*); they have been used to isolate the corresponding nuclear genes (*5, 11*) and study their organization within the genome. Southern blot analysis (*12*) of nuclear DNA from these and several other higher plants have revealed that the rbcS is encoded by a multigene family (*5, 11, 13–15*).

One approach to understanding the tissue-specific, light-dependent expression of rbcS is to construct mutations in putative regulatory regions and study their effects on gene expression (*16*). Such experiments require a system for plant cell transformation. We therefore used tumor-inducing (Ti) plasmid vectors to introduce a pea rbcS gene into petunia cells (*17*). During transformation of susceptible plant cells by virulent strains of *Agrobacterium tumefaciens*, a segment of the Ti plasmid, called T (transferred) DNA, is inserted and stably incorporated into the nuclear DNA (*18*). We now show that a pea rbcS gene (pPS4.0) is expressed after its transfer into petunia cells. The expression of the pea rbcS gene is under the transcriptional control of its own promoter and is regulated by light in a manner similar to that observed in pea leaves. Messenger RNA (mRNA) transcripts from the transferred pea gene are translated to yield mature rbcS polypeptides that assemble with endogenous petunia rbcL polypeptides to form heterologous holoenzymes.

Transformation of petunia protoplasts. The plasmid pMON145 intermediate vector is a variant of the previously described pMON120 (*17*). It contains the 1.6-kilobase (kb) Pvu II to Pvu I segment of pBR322 that carries the origin of replication and *bom* site (*19*) joined to a 2.7-kb segment Cla I to Eco RI of Tn7 (Fig. 1), which contains the spectinomycin-streptomycin resistance determinant

Richard Broglie, Gloria Coruzzi, and Nam-Hai Chua are in the Laboratory of Plant Molecular Biology, The Rockefeller University, New York 10021-6399. Robert T. Fraley, Stephen G. Rogers, Robert B. Horsch, Jeanne G. Niedermeyer, Cynthia L. Fink, and Jeffery S. Flick are on the staff of the Monsanto Company, 800 Lindbergh Boulevard, St. Louis, Missouri 63167. Please address correspondence to N.-H. Chua.