Biochem. J. (1987) 246, 227–232 (Printed in Great Britain)

The binding of the cyclic AMP receptor protein to synthetic DNA sites containing permutations in the consensus sequence TGTGA

Claudia JANSEN, Angela M. GRONENBORN* and G. Marius CLORE
Max-Planck Institut fur Biochemie, D-8033 Martinsried bei München, Federal Republic of Germany

The binding of the cyclic AMP receptor protein (CRP) to symmetrical synthetic DNA-binding sites was investigated with a gel-retardation assay. A set of ten different sequences was employed, comprising all base permutations at positions 2, 4, and 5 of the consensus sequence 5′(TGTGA)3′. We show that: (i) CRP has a higher affinity for the completely symmetrical site than towards the *lac* wild-type site; (ii) base substitutions at position 2 lead to either a complete loss of specific CRP binding (G→C), a reduction in specific CRP binding (G→A) or only marginal effects on specific CRP binding (G→T); (iii) changes at position 4 abolish (G→C; G→A) or reduce (G→T) specific CRP binding; and (iv) base permutations at position 5 reduce specific CRP binding, but never completely abolish it. Thus position 4, and to a lesser extent position 2, in the DNA consensus sequence are the most crucial ones for specific binding by CRP.

INTRODUCTION

Catabolite-sensitive operons in *Escherichia coli* are activated by the cyclic AMP receptor protein (CRP), which, in the presence of cyclic AMP, binds to specific sites near the promoter. CRP-binding sites are also found at other promoters, with CRP acting either as an activator or repressor of transcription. A large number of these sites has been sequenced and a consensus sequence has been proposed (for the latest review, see Busby, 1986). This contains the motif 5′(TGTGA)3′, which is thought to be the crucial point of interaction with the protein (Ebright *et al.*, 1984*a*,*b*). Most CRP sites contain one TGTGA motif and a second sequence that is symmetrically related to the first motif, exhibiting various degrees of symmetry. The affinity of CRP for the different sites may be correlated with the degree of symmetry in the site, with CRP binding more strongly to the most symmetrical *lac* site, followed by the *mal* and *gal* sites (Kolb *et al.*, 1983*a*,*b*).

Detailed models for the interaction of CRP with the *lac* site have been proposed (Weber & Steitz, 1984; Ebright *et al.*, 1984*b*), involving specific contacts with the GC base-pair at position 2, the GC base-pair at position 4 and the AT base-pair at position 5 in the TGTGA motif. These positions have also been implicated in *gal* activation by CRP from random-mutagenesis studies (Busby & Dreyfus, 1983). In order to assess the relative importance of particular bases in the TGTGA motif, we have investigated the binding of CRP to a whole array of synthetic-DNA binding sites. First, we show that a completely symmetric site binds better to CRP than the best natural site, namely the *lac* site. In addition, we tested all possible sequence permutations at positions 2, 4 and 5 in the TGTGA motif. One permutation at position 2 leads to complete loss of specific binding, a second to a noticeable reduction, whereas a third shows only a small reduction with respect to the binding of CRP to the symmetrical parent sequence. Two permutations at position 4 of the TGTGA motif lead to loss of specific binding, whereas the third shows a reduction. All changes at position 5 lead to various degrees of reduction in specific binding of CRP to DNA.

MATERIALS AND METHODS

Bacterial strains and plasmids

E. coli strain JM83 and the vector pUC 19 was used for all cloning work (Yanisch-Perron *et al.*, 1985).

Preparation of plasmid DNA, restriction-endonuclease cleavage, isolation and labelling of fragments, ligation and transformation were carried out as described by Maniatis *et al.* (1982).

CRP purification

CRP was purified from an overproducing *E. coli* strain harbouring the heat-inducible plasmid pPLcCRP1 which carries the *crp* gene under transcriptional control of the λ promoter P_L (Gronenborn & Clore, 1986). It was > 98% pure as judged by SDS/polyacrylamide-gel electrophoresis. Its concentration was determined spectrophotometrically by using a value for ϵ_{278} of 4.1×10^4 M^{-1}·cm^{-1} for the dimer (Takahashi *et al.*, 1982).

Oligonucleotides

34-bases-long oligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer. They were purified by reverse-phase h.p.l.c. and subsequent preparative gel electrophoresis on 20% polyacrylamide gels containing 8 M-urea. Their sequence was verified after cloning by the dideoxy method (Sanger *et al.*, 1977) in the plasmid (Chen & Seeburg, 1985).

Polyacrylamide-gel electrophoresis of DNA–protein complexes

This was essentially carried out as described previously (Garner & Revzin, 1981; Kolb *et al.*, 1983*b*). The DNA fragments used were either a 200-bp endonuclease-*Hin*fI/*Hae*III fragment carrying the *lac* wild-type (w-t)

Abbreviations used: CRP, cyclic AMP receptor protein of *Escherichia coli*; w-t, wild-type; DTT, dithiothreitol; BSA, bovine serum albumin.

* To whom correspondence and reprint requests should be addressed.

region or 85-bp *Hin*dIII/*Eco*RI fragments derived from the pUC 19 clones carrying the synthetic oligonucleotides. Radioactive end-labelling was achieved with Klenow polymerase and [α-^{32}P]ATP. The approximate concentration of DNA fragments used was 10^{-10} M. Samples containing various amounts of CRP were incubated for 30 min at room temperature in 40 mM-Tris (pH 8)/10 mM-$MgCl_2$/10 mM-KCl/1 mM-DTT/BSA-(0.1 mg/ml)/0.2 mM-cyclic AMP. Electrophoresis was performed on 12% polyacrylamide gels for 2 h in 30 mM-Tris/90 mM-borate/2.5 mM-EDTA/0.2 mM-cyclic AMP. The gels were autoradiographed at −70 °C with intensifying screens for 12 h.

RESULTS

Fig. 1 shows a comparison of the *lac* w-t sequence with the sequences of all ten oligonucleotides used in the present study. The parent sequence, Sym Lac 1, is a completely symmetrical sequence, which differs from the *lac* w-t sequence by a C→T change in the space between the two TGTGA motifs and by an altered sequence at the 3′ end of the pentanucleotide motif. This particular choice of sequence was stimulated by the observed correlation between binding strength and symmetry (Kolb *et al.*, 1983*b*). Thus a completely symmetrical sequence should in principle bind CRP better than the best natural sequence, a hypothesis which we set out to test. Furthermore, because of the palindromic nature of the designed oligonucleotide, each mutant binding site required the synthesis of only one oligonucleotide. The sequence at the 5′ and 3′ ends was chosen in such a way that, upon annealing, *Bam*H1 ends were created for ease of cloning. Each oligonucleotide was inserted into the *Bam*H1 site of pUC 19 (Yanisch-Perron *et al.*, 1985) and the sequence was verified by plasmid DNA sequencing (Chen & Seeburg, 1985). Binding experiments were performed with 85 bp-long *Hin*dIII/*Eco*RI fragments of the various pUC19 clones, which carry the CRP-binding site on the synthetic oligonucleotide approximately in the middle.

Fig. 2(*a*) shows a binding experiment using a 200-bp *lac* w-t fragment (lanes a–f) and the 85 bp fragment carrying

W-t Lac	CATTAATGTGAGTTAGCTCACTCATTAG GTAATTACACTCAATCGAGTGAGTAATC
Sym Lac 1	GATCCAATTAATGTGAGCTAGCTCACATTAATTG GTTAATTACACTCGATCGAGTGTAATTAACCTAG
Sym Lac 2	GATCCAATTAATCTGAGCTAGCTCAGATTAATTG GTTAATTAGACTCGATCGAGTCTAATTAACCTAG
Sym Lac 3	GATCCAATTAATTTGAGCTAGCTCAAATTAATTG GTTAATTAAACTCGATCGAGTTTAATTAACCTAG
Sym Lac 4	GATCCAATTAATATGAGCTAGCTCATATTAATTG GTTAATTATACTCGATCGAGTATAATTAACCTAG
Sym Lac 5	GATCCAATTAATGTCAGCTAGCTGACATTAATTG GTTAATTACAGTCGATCGACTGTAATTAACCTAG
Sym Lac 6	GATCCAATTAATGTTAGCTAGCTAACATTAATTG GTTAATTACAATCGATCGATTGTAATTAACCTAG
Sym Lac 7	GATCCAATTAATGTAAGCTAGCTTACATTAATTG GTTAATTACATTCGATCGAATGTAATTAACCTAG
Sym Lac 8	GATCCAATTAATGTGCGCTAGCGCACATTAATTG GTTAATTACACGCGATCGCGTGTAATTAACCTAG
Sym Lac 9	GATCCAATTAATGTGTGCTAGCACACATTAATTG GTTAATTACACACGATCGTGTGTAATTAACCTAG
Sym Lac 10	GATCCAATTAATGTGGGCTAGCCCACATTAATTG GTTAATTACACCCGATCGGGTGTAATTAACCTAG

Fig. 1. Base sequences of the CRP binding sites used in the present study

The site of permutation within the TGTGA motif is marked by an asterisk. Each 34-mer was synthesized, cloned into the endonuclease-*Bam*H1 site of pUC 19 and excised in the form of 85-bp-long *Hin*dIII/*Eco*RI fragments. The latter were used in the DNA-binding assays.

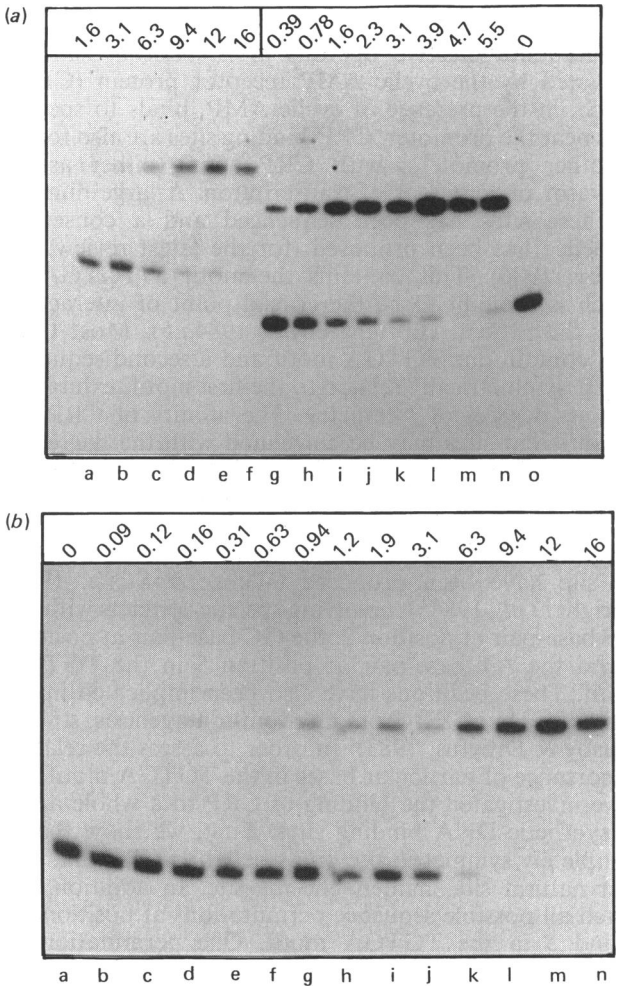

Fig. 2. (*a*) Comparison of CRP binding to a 200 bp *lac* w-t fragment (left-hand side) and a synthetic symmetrical binding site, Sym Lac 1 (right-hand side), and (*b*) a complete binding experiment for Sym Lac 1

The concentration of CRP (nM) in the binding reaction is indicated at the top of each lane.

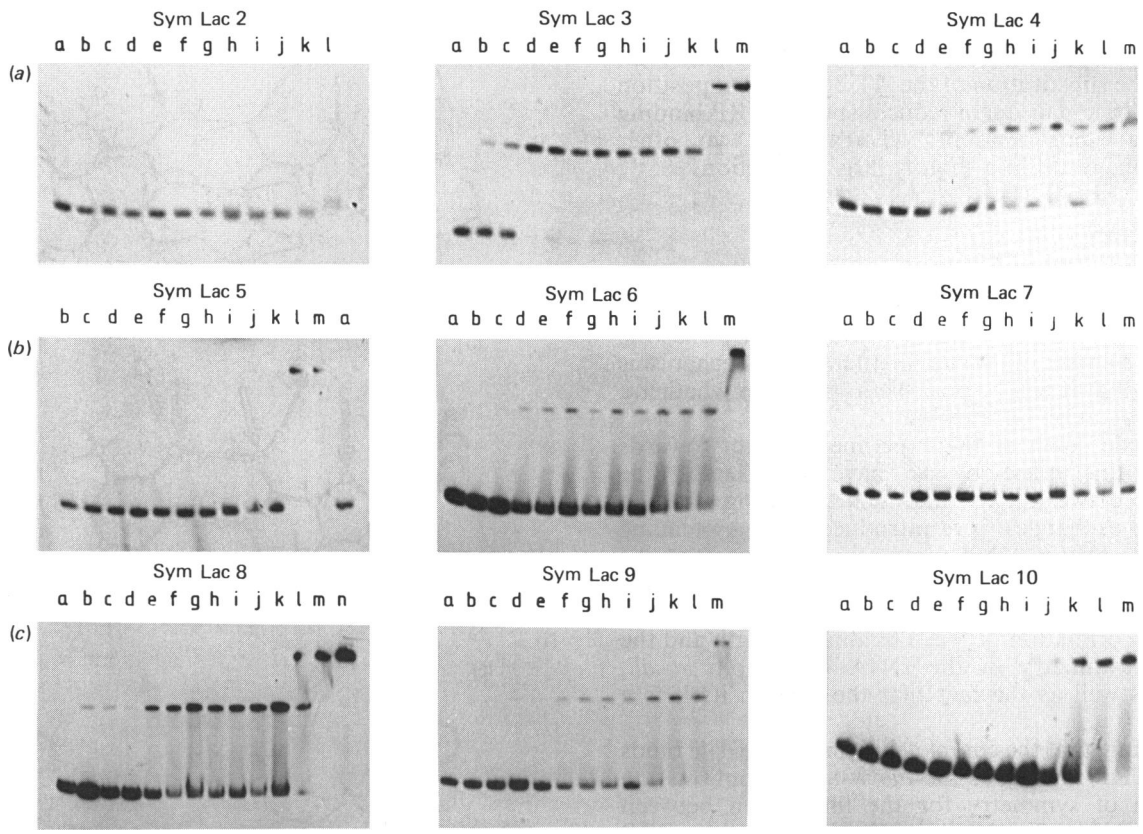

Fig. 3. Gel-retardation assays for all permutated CRP-binding sites

Lanes a, b, c, d, e, f, g, h, i, j, k, l, m, n correspond to 0 M-, 3.1×10^{-9} M-, 6.2×10^{-9} M-, 1.2×10^{-8} M-, 1.9×10^{-8} M-, 2.5×10^{-8} M-, 3.1×10^{-8} M-, 3.8×10^{-8} M-, 4.4×10^{-8} M-, 1.6×10^{-7} M-, 3.1×10^{-7} M-, 6.2×10^{-7} M-, 1.2×10^{-6} M-, 1.9×10^{-6} M-CRP respectively. (*a*) G at position 2 substituted by C, T and A; (*b*) base changes at position 4; (*c*) base changes at position 5. For further details, see the text.

Sym Lac 1 (lanes g–n). It is apparent that CRP binds better to the 85 bp fragment than to the 200 bp fragment, as at equal amounts of protein (lanes b and k) only a very small amount of protein–DNA complex is present in the case of the *lac* w-t sequence, whereas, in case of the Sym Lac 1 sequence, hardly any uncomplexed DNA is detectable. Thus, as predicted, CRP binds better to the more symmetrical site than to the *lac* w-t site.

Fig. 2(*b*) shows a complete binding experiment using the Sym Lac 1 fragment. By measuring either the amount of specific DNA–protein complex formed or the residual amount of free DNA, it is possible to estimate a binding constant. This is approx. 10^9 M^{-1}. Under the same experimental conditions the binding constant for specific CRP binding to the 200 bp *lac* w-t fragment is around $(2–3) \times 10^8$ M^{-1}.

Fig. 3 shows similar binding experiments using fragments that carry changes in the TGTGA motif. The first three panels (*a*) show gel-binding assays for sequences in which the G at position 2 has been substituted by C, T and A. As is easily observed, binding of CRP to these sites is reduced in all cases compared with the TGTGA containing Sym Lac 1 sequence. The G→C change abolishes specific binding completely (Sym Lac 2), whereas the G→T change (Sym Lac 3) leads to only a small reduction, so that the apparent binding constant lies in the range of that found for the *lac* w-t fragment. The last permutation of position 2, G→A (Sym Lac 4), leads to gels in which the 1:1 CRP–DNA complex only becomes apparent at CRP concentrations greater than 1×10^{-8} M. Thus the binding constant has to be reduced by a factor of ~ 50 compared with Sym Lac 1.

The next three panels (*b*) show binding gels using DNA sequences which contain base changes at position 4 of the TGTGA sequence. In the case of the G→C change (Sym Lac 5), no 1:1 complex can be observed; at high CRP concentrations, however, non-specific binding to any sequence on the fragment results in large protein–DNA aggregates which are visible in the slots of the gel. CRP binding to Sym Lac 6 (G→T) shows 1:1 complexes at similar CRP concentrations to those found for the analogous change at position 2 (Sym Lac 3), so that this change results in only a small reduction in the binding constant. For Sym Lac 7 (G→A), on the other hand, no binding of CRP is detectable at all.

The last three panels (*c*) show the effects on the specific binding of CRP to DNA which result from changes in position 5 of the TGTGA sequence. All permutations lead to some loss of binding, with the A→C change (Sym Lac 8) resulting in only a small reduction and the A→T (Sym Lac 9) and A→G (Sym Lac 10) changes in somewhat larger loss of binding.

These results may be summarized as follows: changing the GC base-pair at either position 2 or 4 in the TGTGA motif of the synthetic palindromic DNA site to a TA

base-pair reduces specific CRP binding by only a small amount; all other changes at these two positions result either in complete loss of, or a large reduction in, binding. The substitution of the AT base-pair at position 5 by a CG base-pair again reduces specific CRP binding by only a small amount, whereas the two other permutations result in a slightly larger reduction.

DISCUSSION

The ability of a protein to recognize specifically a particular base sequence of DNA is of fundamental importance in molecular biology, and it is only once the details of a number of individual interaction mechanisms are known that it will be possible to ascertain whether a general recognition code exists.

To provide some of the experimental data towards answering this question, we have investigated the interaction between CRP and DNA binding sites, in which base exchanges were introduced in a systematic fashion. The chosen DNA sequences were based on the *lac* w-t sequence, introducing perfect symmetry into the site. The principle reason for this particular choice was the known correlation between binding strength and the degree of symmetry in the DNA site (Kolb *et al.*, 1983*a*,*b*) as well as the fact that the protein itself is a dimer.

Our finding, that the completely symmetrical site binds better to CRP than does the *lac* w-t site, confirms the importance of symmetry for the interaction between CRP and DNA. Furthermore, it suggests that CRP exhibits twofold symmetry in the complex, although the crystal structure of the cyclic AMP·CRP complex is asymmetric, with 'open' and 'closed' subunits (McKay *et al.*, 1982). This asymmetry, however, may solely be due to crystal-packing forces, as the 'open' and 'closed' subunits are in close protein–protein contact with a second CRP molecule in the crystal lattice (McKay & Steitz, 1981). In addition, one can conclude, on the basis of the observed strong binding, that the 34-mer symmetric site contains all the necessary distal elements contributing to the binding interaction, in particular bases 12 and 14 nucleotides distant from the dyad axis, which were found to be important for binding by Liu-Johnson *et al.* (1986).

The relative importance of individual bases in the TGTGA motif can be deduced from the relative binding affinities for the different permutated sites. The observation that the affinity of CRP for the synthetic symmetrical site is increased by a factor of approx. 3–5 relative to the *lac* w-t site also points to the importance of the first T in the pentanucleotide consensus sequence. This change, however, is not as substantial as one might expect if an additional hydrogen bond was formed, and may be due to either a sequence-dependent minor change in the DNA structure leading to a better interaction with the sugar-phosphate backbone or to a hydrophobic interaction between the protein and the thymine methyl group.

A general hydrogen-bonding scheme for base-pairs with amino acids in the major groove has been proposed (Seeman *et al.*, 1976) and will be used in the discussion below. Fig. 4 illustrates the possible hydrogen-bonding interactions involving the major groove edges of the bases.

A GC or CG base-pair can interact with amino acid side chains protruding into the major groove, either by donating a hydrogen bond from the 4-NH_2 group of cytosine or by accepting hydrogen bonds at the 6-keto or N7 position of guanine. Substitution of the GC base-pair by a TA base-pair at either position 2 or 4 in the consensus sequence leads to only a small perturbation in the binding by CRP. This result can be explained on the basis that two of the possible hydrogen-bonding positions are only marginally shifted between GC and TA base-pairs. Thus the 4-NH_2 group of cytosine is in a similar position to the 6-NH_2 group of adenine [shifted by 0.12 nm (1.2 Å)], whereas the 6-keto group of guanine is in a similar location to the 4-keto group of thymine [shifted by 0.1 nm (1 Å)], as illustrated in Fig. 4(*a*). It therefore seems likely that either one or both of these two groups interact with the protein. The N7 of guanine, on the other hand, is less likely to be involved in the interaction, since the methyl group of thymine in a TA base-pair is located in the equivalent position and would thus be expected to result in steric hindrance.

For the GC-to-AT base-pair change at either position

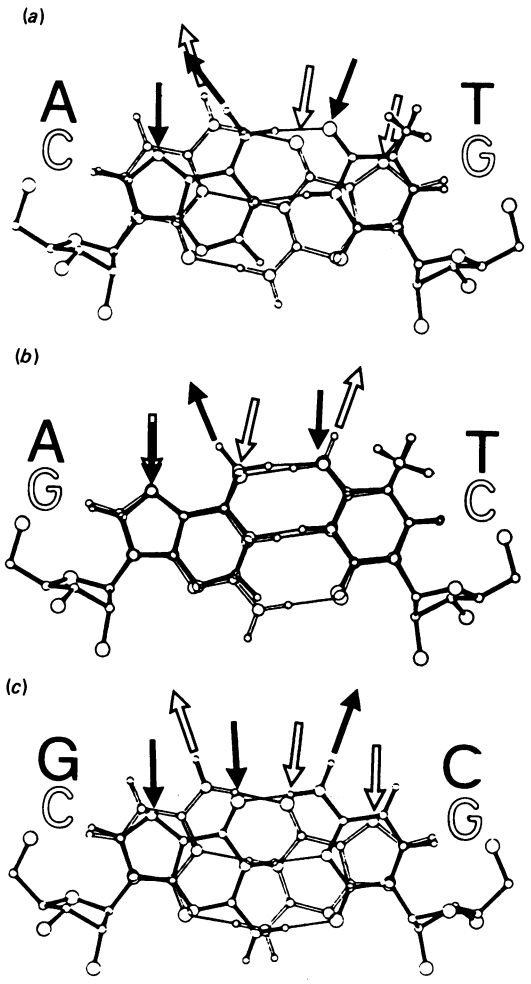

Fig. 4. Possible hydrogen-bonding sites for base-pairs in the major groove (adapted from Seeman *et al.*, 1976)

To illustrate the different positions of hydrogen-bond acceptor and donor sites for alternative base-pairs, different combinations of base-pairs are superimposed in (*a*), (*b*) and (*c*). Arrows indicate acceptor/donor sites for protein interaction.

2 or 4 the specific binding of CRP is noticeably reduced ($\sim$ 40 fold), equivalent to a loss in interaction energy of approx. 9.2 kJ (2.2 kcal). This suggests the loss of two hydrogen bonds for this base-pair change, and, as the TGTGA motif occurs twice in the fragment, this represents the loss of one hydrogen bond per TGTGA sequence (see Fig. 4*b*). The GC-to-CG base-pair change (see Fig. 4*c*) finally leads to a complete loss of specific binding, possibly arising from the loss of the same hydrogen bonds and an additional reduction in non-hydrogen-bonding interactions.

The amino acid proposed to interact with the GC pair at position 2 is Arg-180 (Ebright *et al.*, 1984*b*; Weber & Steitz, 1984), donating two hydrogen bonds to the guanine in a bidentate fashion, as illustrated in Fig. 5. If this proposal is correct, then the hydrogen bond between the N7 of guanine and the amino acid side chain has to be substantially weaker than the one formed with the 6-keto group, as the loss of this hydrogen bond in the GC-to-TA base-pair change reduces the binding only by a factor of 3–5, whereas loss of the hydrogen bond accepted by the oxygen reduces the binding $\sim$ 40 fold.

The GC base-pair at position 4 is in contact with Glu-181, as evidenced by genetic data (Ebright *et al.*, 1984*a*), and donates a hydrogen bond via the 4-NH_2 group of cytosine. This is shown schematically in Fig. 5. Here too, the change to a TA base-pair results in only a 0.12 nm (1.2 Å) shift of the amino hydrogen atom if one compares the 4-NH_2 group of cytosine with the 6-NH_2 group of adenine (see Fig. 4*a*), consistent with a reduction in binding, but not a complete loss. The latter is the case for the two other possible permutations. Incidentally, the change from a GC to an AT base-pair is the equivalent in this symmetrical sequence to the L8 and L29 mutations in the *lac* w-t sequence, for which a reduction in specific CRP binding affinity has also been observed (Kolb *et al.*, 1983*a*). As one of the two TGTGA motifs is not altered in the *lac* case, a 1:1 CRP–DNA complex can still be observed in the gel assays for the L8 and L29 fragments (Kolb *et al.*, 1983*a*). In our case, however, no intact TGTGA is available for binding, owing to the symmetrically introduced change, and this is manifested by the absence of any specific binding for this altered sequence.

Position 5 of the consensus motif is supposed to interact by accepting a hydrogen bond from Arg-185 at the N7 of the adenine. This is probably not the case, since changing the AT base-pair to a CG base-pair, which does not carry a hydrogen-bond acceptor at a position equivalent to the N7 of adenine (see Fig. 4*a*), results in a smaller reduction in binding than the change to a GC pair (see Fig. 4*c*), which possesses an N7 in the same place. For this position none of the permutations resulted in a complete loss of specific binding, suggesting that the proposed model may not be correct in this instance. Corroborating evidence towards this suggestion comes from DNA-binding studies using mutant proteins in which Arg-185 was changed to either Lys or Leu. In both cases transcription activity *in vivo* and specific binding to the *lac* site was similar to that of w-t CRP (Gent *et al.*, 1987; A. M. Gronenborn, unpublished work). Thus it is clear that Arg-185 cannot play an important role in the specific interaction, and it could be that the influence of the AT base-pair at position 5 on the binding of CRP is mainly due to changes in the sugar-phosphate backbone structure and its interactions.

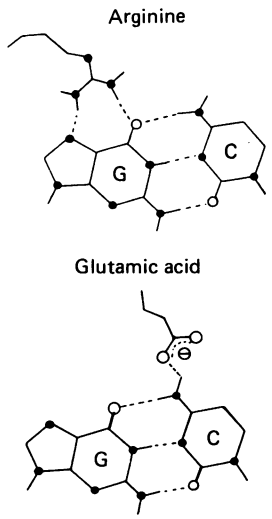

Fig. 5. Proposed interaction between guanosine and arginine and between cytosine and glutamic acid in a GC base-pair

Alternatively, a specific interaction between the AT base-pair and another amino acid of CRP may occur, involving either the 6-NH_2 of adenine or the 4-keto group of thymine, as exchange to a CG base-pair reduces the binding by only a relatively small amount. Here too, both the amino and keto functions are only shifted by $\sim$ 0.1 nm (1 Å), so that similar interactions could occur (see Fig. 4*a*). This would be analogous to the situation found at base-pair 2. Inspection of the protein sequence, however, does not reveal any amino acid, apart from Arg-185, that could be involved in such specific hydrogen-bonding. Alternatively, a hydrophobic interaction involving the thymine methyl group or a non-specific interaction between Arg-185 and the DNA-phosphate backbone may be important. Indeed, ethylation interference experiments have shown that the phosphate at the 5′ side of the thymine in the AT base-pair is important for CRP binding (Majors, 1979).

From these results one can draw the following conclusions. (*a*) Position 4 of the TGTGA motif is the most important one for determining specific binding, as only one substituted sequence remains able to interact specifically with CRP, albeit with reduced binding affinity. (*b*) Position 2 shows loss of specific binding for one permutation (G→C), which also indicates a strong interaction via hydrogen-bonding. (*c*) All changes in position 5 lead to some, but never complete, loss of binding, indicative of changes in non-specific interactions, most likely involving the phosphate backbone.

Thus the proposed model for the specific interaction of CRP with DNA has to be amended as follows: hydrogen bonds occur between Arg-180 and the 6-keto group of guanine in position 2 and between Glu-181 and the 4-NH_2 group of cytosine in position 4, but not between Arg-185 and the adenine in position 5 of the TGTGA consensus sequence.

This work was supported by the Max-Planck Gesellschaft and Grant no. Gr658/3-1 of the Deutsche Forschungsgemeinschaft (to G.M.C. and A.M.G.). C.J. thanks the Max-Planck Gesellschaft for a pre-doctoral fellowship. The skilful technical assistance of Mr. S. Gärtner is gratefully acknowledged.

REFERENCES

Busby, S. (1986) in Regulation of Gene Expression: 25 Years On (Booth, I. R. & Higgins, C. F., eds.), pp. 51–77, Cambridge University Press, Cambridge

Busby, S. & Dreyfus, S. (1983) Gene **21**, 121–131

Chen, E. Y. & Seeburg, P. H. (1985) DNA **4**, 165–170

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984*a*) Nature (London) **311**, 232–235

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984*b*) Proc. Natl. Acad. Sci. U.S.A. **81**, 7274–7278

Garner, M. & Revzin, A. (1981) Nucleic Acids Res. **9**, 3047–3060

Gent, M. E., Gronenborn, A. M., Davies, R. W. & Clore, G. M. (1987) Biochem. J. **242**, 645–653

Gronenborn, A. M. & Clore, G. M. (1986) Biochem. J. **236**, 643–649

Kolb, A., Spassky, A., Chapon, G., Blazy, B. & Buc, H. (1983*a*) Nucleic Acids Res. **22**, 7833–7852

Kolb, A., Busby, B., Herbert, M., Kotlarz, D. & Buc, H. (1983*b*) EMBO J. **2**, 217–222

Liu-Johnson, H.-N., Gartenberg, M. R. & Crothers, D. M. (1986) Cell **47**, 995–1005

Majors, J. (1979) Ph. D. Thesis, Harvard University

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

McKay, D. B. & Steitz, T. A. (1981) Nature (London) **290**, 744–749

McKay, D. B., Weber, I. T. & Steitz, T. A. (1982) J. Biol. Chem. **257**, 9518–9524

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 804–808

Takahashi, M., Gronenborn, A. M., Clore, G. M., Blazy, B. & Baudras, A. (1982) FEBS Lett. **139**, 37–40

Weber, I. T. & Steitz, T. A. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 3973–3977

Yanisch-Perron, C., Viera, J. & Messing, J. (1985) Gene **33**, 103–119

Received 2 February 1987/27 April 1987; accepted 21 May 1987