Volume 13 Number 16 1985 Nucleic Acids Research

The primary structure of the DeoR repressor from *Escherichia coli* K-12

Poul Valentin-Hansen*, Peter Højrup and Steven Short[1]

Department of Molecular Biology, Odense University, DK-5230 Odense M, Denmark, and
[1]Department of Microbiology, Wellcome Research Laboratories, Research Triangle Park, NC, USA

Received 23 May 1985; Revised and Accepted 5 August 1985

ABSTRACT

The nucleotide sequence of the deoR gene of E. coli, which codes for the DeoR repressor, has been determined. This gene codes for a polypeptide that is 252 amino acids residues in length. Computer-assisted analysis of the nucleotide sequence strongly suggests that the DNA binding domain of the DeoR repressor is located in the N-terminal part of the protein. After the coding region there is a dyad symmetry similar to a palindromic unit present outside many structural genes on the E. coli chromosome.

INTRODUCTION

Expression of the majority of genes involved in uptake and catabolism of nucleosides and deoxyribonucleosides in E. coli is regulated negatively by one or two regulatory proteins, the DeoR and CytR repressors (reviews of earlier work can be found in Refs. 1 and 2). The promoter regions for these genes can be classified in three regulatory groups: i) DeoR regulated, ii) CytR regulated and iii) CytR and DeoR regulated. In addition, the cAMP/CRP complex is required for activation of all the CytR controlled transcriptional units. The DeoR repressor is known to regulate the deoCABD genes, encoding deoxy- and ribonucleoside catabolizing enzymes (3-5), and the nupG and tsx genes which encode a transport protein and a pore-forming protein, respectively (6,7).

It has been shown both in vivo and in vitro that the deo operon is controlled by two promoters, P1 and P2, located 600 bp apart in front of the operon. Initiation of transcription from both promoter regions is negatively controlled by the DeoR repressor, and it is suggested that the DeoR recognition sites in the promoter regions overlap the initiation site for transcription (8-10) (Fig. 1).

The deoR gene maps at 18.7 min on the E coli chromosome close to the cmlA gene and has recently been located on a 2.5 kilobase SphI-BamHI fragment; a SalI and ClaI site was localized within either the structural

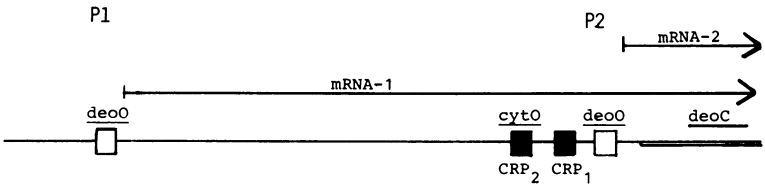

Figure 1. Schematic map of the deoP1 and P2 regulatory region. At P1 and P2 transcription is repressed by the DeoR repressor and induced by deoxyribose-5-P; at P2 transcription is also repressed by the CytR repressor (the inducer being cytidine or adenosine) and depends on cAMP and CRP. deoO is the sites recognized by DeoR; CRP_1 and CRP_2 the cAMP/CRP targets. The CytR operator (cytO) most likely overlaps the CRP_2 target.

gene or in its promoter region, since SalI or ClaI deletions failed to direct the synthesis of active repressor (11). Using this information we have determined the nucleotide sequence of the deoR gene and its promoter region.

MATERIALS AND METHODS

Bacterial strains and plasmids

The relevant genotypes of the E. coli strains used were: Sø928 (Δdeo-11,Δlac); Sø 3135 (Δdeo,Δlac,Φdeo lac,deoR$^-$). The plasmids used were pSS344 (11), pUC13 (12).

Strain Sø3135 was used for cloning of the deoR gene. In this strain the expression of the lac genes is controlled by the deoP1 and P2 promoter regions. The lac phenotype was monitored on lactose-McConkey indicator plates. Cells containing deoR$^+$ plasmids appeared white on these plates.

Enzymes and chemicals

Restriction endonucleases were purchased from commercial suppliers (Boehringer, New England Biolabs). Conditions for restriction endonuclease digestions were those recommended by the commercial suppliers. T4 DNA ligase, Bal-31 nuclease, E. coli DNA polymerase I (large fragment), bacterial alkaline phosphatase, S1 nuclease and T4 polynucleotide kinase were obtained from BRL, and γ ^{32}P- and α ^{32}P-labeled nucleotide triphosphates from New England.

Construction of deoR recombinant plasmids

Plasmid pVH335 was constructed in the following way: pSS344 DNA was digested with SphI and treated with Bal-31 nuclease. Samples containing deletions of approximately 2500 bp were ligated with T4 DNA ligase and

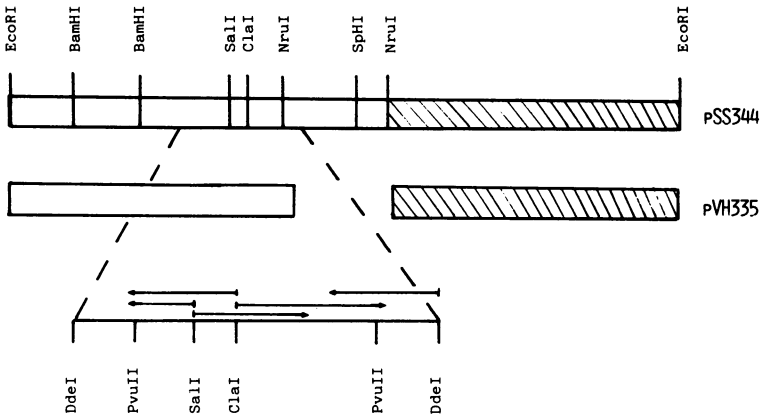

Figure 2. Restriction maps of deoR$^+$ plasmids. The hatched segments represent pBR322 DNA, and the open segments represent cloned DNA. The deleted DNA in pVH335 is indicated by a gap. Shown at the bottom is the DdeI fragment cloned into pUC13, along with the restriction sites used for DNA sequencing. The length of the arrows indicates the extent of the sequence determined.

transformed into Sø3135. Cells harbouring deoR$^+$ plasmids were selected as white (lac$^-$)Ap$^+$ colonies using lactose-McConkey indicator agar containing 100 µg/ml ampicillin.

Plasmids pVH336 and pVH337 were constructed as follows: A total of 2 µg of DdeI restricted pSS344 was treated with S_1 nuclease and blunt-end ligated with SmaI restricted plasmid pUC13 and transformed into Sø3135. Ampicillin-resistant deoR$^+$ transformants were selected as described for pVH335. The orientation of the inserted DdeI fragment was mapped with respect to the vehicle, being: lacP-(deoR-P)-deoR for pVH336 and lacP-deoR--(deoR-P) for pVH337.

RESULTS AND DISCUSSION

Sequencing strategy

The structural gene for the DeoR repressor is known from cloning experiments to be contained on a 2.7 kb SphI-BamHI restriction fragment, and a SalI and a ClaI site have been located within the gene or in its promoter region (11), Fig. 2. To determine the location more precisely several deletions were constructed from the unique SphI site in plasmid pSS344 using Bal-31 nuclease. Plasmids containing deletions of the NruI site present in the SphI-BamH1 fragment failed to direct the synthesis of functional DeoR repressor, whereas plasmids with deletions located around

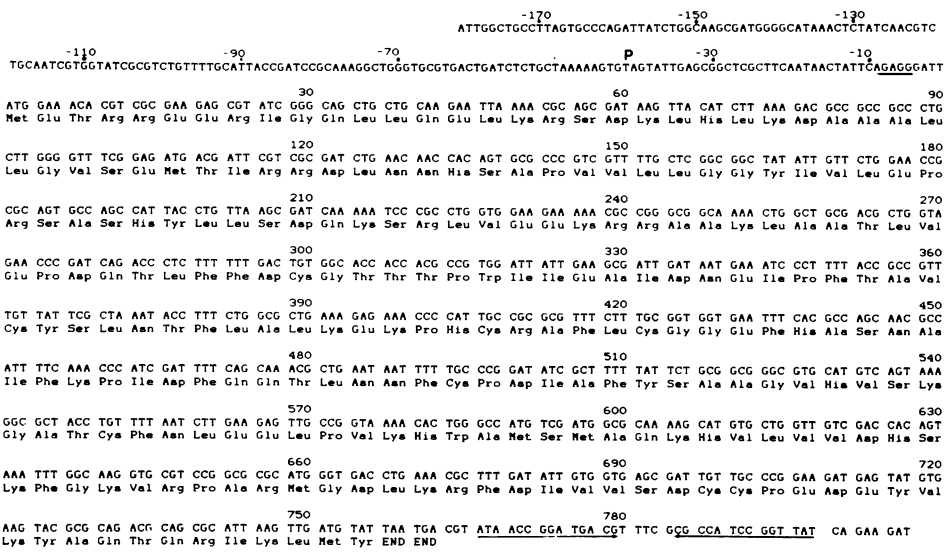

Figure 3. The nucleotide sequence and encoded polypeptide of the deoR gene. The putative promoter is represented by the Pribnow box (P), and the Shine-Dalgarno sequence is underlined.

the DdeI site in front of the NruI site still express active DeoR repressor (plasmid pVH335, Fig. 2). Using this information we sequenced the DNA region from DdeI to PvuII as illustrated in Fig. 3. The DNA sequence was determined of both strands, and sequences at or near fragment junctions were confirmed with overlapping fragments. The complete nucleotide sequence is presented in Fig. 3.

The DNA sequence of the repressor gene

Examination of a reading frame which should start or end in front of the NruI site in plasmid pSS344 revealed that only one could code for a polypeptide of significant length. This potential translated sequence begins at the ATG codon at position 1-3 and extends over both the ClaI and SalI restriction sites to the stop codons TAA-TGA at position 757-762 (Fig. 3). Assuming that the protein is not modified the polypeptide predicted from this sequence would be 252 amino acid residues in length with a molecular weight of 28550 daltons. The predicted size of the subunit of DeoR is consistent with the finding that deoR$^+$ plasmids direct the synthesis of an approximately 27 kd polypeptide in maxicells (not shown).

The reading frame thus established is preceded by a hypothetical ribosome binding site (underlined in Fig. 3) complementary to the 3' end of

16S RNA (13). The ATG codon at position 1-3 is the only likely candidate for the initiation codon, since no other in-frame start codon (TTG, GTG, ATT) is preceded by a Shine-Dalgarno-like sequence.

Recloning of the deoR repressor

To increase the expression of the DeoR repressor and to confirm the reading frame for the repressor gene we recloned the DdeI fragment from pSS344 which spans the sequenced region (Fig. 2). The vector chosen for the cloning experiments was pUC13 (12). This plasmid carries the lac regulatory region and the first part of lacZ in which a polylinker with multiple cloning sites has been inserted.

Whatever its orientation the DdeI fragment from pSS344 when cloned into the SmaI site of pUC13 expresses the repressor gene (plasmid pSS336 and pSS337, See Materials and Methods). However, considerably more repressor is synthesized from pVH336 than from pVH337. Since the expression of the open reading frame in pVH336 is under the control of the lac promoter present in pUC13 we believe that the protein sequence in Fig. 3 represents the polypeptide of DeoR.

DNA sequences flanking the deoR gene

In an attempt to locate the deoR promoter in the sequenced DNA preceding the open reading frame we looked for regions showing homology with the conserved -10 (TATAAT) and -35 regions (TTGACA) of E. coli promoters (14-16). The stretch of DNA which has the highest homology to the promoter concensus sequences is located from bp -62 to -35 with a spacing of 18 bp between the putative Pribnow box (TAGTAT) and the -35 region (GTGACT). Assuming that the mRNA is initiated 4-7 bp after the Pribnow sequence the mRNA should have a leader sequence of about 30 bp.

The putative Shine-Dalgarno sequence (GAGG) for deoR is located very close to the start codon (underlined in Fig. 3). The spacing is only 4 bp, which is unusual (17). This short distance may well contribute to limit the translation efficiency of deoR mRNA.

Inspection of the DNA sequence in the region downstream from the open reading frame reveals the presence of a long dyad symmetry which, once transcribed, could form a stable stem-loop structure with a stem of 14-15 bp (Fig. 3). The stem-loop structure shows high homology to a genetic element present at least several hundred times outside structural genes on the E. coli chromosome (18). It seems likely that this element plays a role in mRNA degradation as well as rearrangements or duplications of genes (18-20).

Table 1 Codon usage and amino acid composition in DeoR
(Percent occurrence of codons and amino acids is given in parentheses)

TTT	12 (4.7)	TCT	1 (0.3)	TAT	5 (1.9)	TGT	4 (1.5)
TTC	1 (0.3)	TCC	1 (0.3)	TAC	2 (0.7)	TGC	4 (1.5)
TTA	3 (1.1)	TCA	0 (0.0)	TAA	0 (0.0)	TGA	0 (0.0)
TTG	3 (1.1)	TCG	3 (1.1)	TAG	0 (0.0)	TGG	2 (0.7)
CTT	4 (1.5)	CCT	1 (0.3)	CAT	5 (1.9)	CGT	4 (1.5)
CTC	2 (0.7)	CCC	4 (1.5)	CAC	4 (1.5)	CGC	10 (3.9)
CTA	1 (0.3)	CCA	0 (0.0)	CAA	4 (1.5)	CGA	0 (0.0)
CTG	14 (5.5)	CCG	6 (2.3)	CAG	5 (1.9)	CGG	1 (0.3)
ATT	8 (3.1)	ACT	0 (0.0)	AAT	5 (1.9)	AGT	4 (1.5)
ATC	4 (1.5)	ACC	6 (2.3)	AAC	3 (1.1)	AGC	5 (1.9)
ATA	0 (0.0)	ACA	1 (0.3)	AAA	12 (4.7)	AGA	0 (0.0)
ATG	6 (2.3)	ACG	5 (1.9)	AAG	5 (1.9)	AGG	0 (0.0)
GTT	5 (1.9)	GCT	3 (1.1)	GAT	10 (3.9)	GGT	3 (1.1)
GTC	3 (1.1)	GCC	8 (3.1)	GAC	4 (1.5)	GGC	6 (2.3)
GTA	2 (0.7)	GCA	1 (0.3)	GAA	12 (4.7)	GGA	0 (0.0)
GTG	7 (2.7)	GCG	11 (4.3)	GAG	5 (1.9)	GGG	2 (0.7)

23	(9.1)	Ala	27 (10.7)	Leu
15	(5.9)	Arg	17 (6.7)	Lys
8	(3.1)	Asn	6 (2.3)	Met
14	(5.5)	Asp	13 (5.1)	Phe
8	(3.1)	Cys	11 (4.3)	Pro
9	(3.5)	Gln	14 (5.5)	Ser
17	(6.7)	Glu	12 (4.7)	Thr
11	(4.3)	Gly	2 (0.7)	Trp
9	(3.5)	His	7 (2.7)	Tyr
12	(4.7)	Ile	17 (6.7)	Val

Codon usage

Analysis of codon usage has revealed that genes for proteins which are abundant in the cell use a narrow set of codons, whereas proteins present in trace amounts, such as repressors, have a broader spectrum of codons (21-23). Thus 22 codons are rarely used in major proteins (<10%) and are called "rare codons". In the deoR gene the rare codons occur with a frequency of 22% similar to that observed for several other repressor genes. In particular the "weak" leu and ser codons have a high preference in the deoR gene where they amount to 50% of all leu and ser codons used (see Table 1).

Similarity to DNA binding proteins

Many regulatory proteins seem to recognize their targets using a common mechanism. The three-dimensional structure of the cro- and λ-repressor of phage lambda and the CAP protein of E. coli have been determined and all three proteins seem to have arrived at a similar solution for recognizing and binding to the DNA helix (24-27). Each protein contains a helix-

λcro	Phe_{14}Gly Gln Thr Lys Thr Ala Lys Asp Leu Gly Val Tyr Gln Ser Ala Ile Asn Lys Ala Ile His Ala
λcI	Leu_{31}Ser Gln Glu Ser Val Ala Asp Lys Met Gly Met Gly Gln Ser Gly Val Gly Ala Leu Phe Asn Gly
P22cI	Gln Arg Lys Val Ala Asp Ala Leu Gly Ile Asn Glu Ser Gln Ile Ser Arg Trp Lys Gly
lacI	Val_4 Thr Leu Tyr Asp Val Ala Glu Tyr Ala Gly Val Ser Tyr Gln Thr Val Ser Arg Val Val Asn Gln
galR	Ala_2 Thr Ile Lys Asp Val Ala Arg Leu Ala Gly Val Ser Val Ala Thr Val Ser Arg Val Ile Asn Asn
deoR	Leu_{22}His Leu Lys Asp Ala Ala Ala Leu Leu Gly Val Ser Glu Met Thr Ile Arg Arg Asp Leu Asn Asn

Figure 4. Comparison of the NH_2-terminal amino acid sequence of the DeoR repressor with those of five other DNA-binding proteins, aligned as suggested by Sauer et. al. (28). Residues in DeoR which are found in identical positions in the other repressors are underlined.

-turn-helix sequence with one of the α-helices able to fit into the major groove of right-handed B-DNA. Comparison of the amino acid sequences of about 20 prokaryotic regulatory proteins has revealed that regions are present in the N- or C-terminal part of these proteins which are homologous to the DNA-binding domain of CAP and the repressor and cro proteins of phage lambda (28-30).

It was, therefore, of interest to compare the amino acid sequences of the DeoR repressor with those of other DNA-binding proteins. In Fig. 4 the DNA-binding region of 5 repressors (all regions are located close to the N-terminal) are aligned to give maximal homology. In DeoR only a single region, starting 21 residues from the N-terminal, shows strong homology to the DNA-binding regions. This part of DeoR is shown as line 6 in Fig. 4.

16 residues in this region of DeoR can be found in positions common to one or more of the other five repressors (underlined in Fig. 4). The homology between the aligned parts of the six repressors varies between 13 and 61% (Table 2) and it can be seen that DeoR in this region is more closely related to the lac and gal repressors than to the other proteins.

The helix-turn-helix secondary structure found in λ-cro and λ-repressor seems to be an invariant feature of the DNA-binding proteins. Three positions are constant in all the proteins sequenced so far (Ala-28, Gly-32 and Ile/Val-38, DeoR numbering) and which are all found in DeoR. The three invariant residues seem to be important in determining the turn and the angle that the helices have to each other (28).

Secondary structure predictions using the methods of Garnier et al. (31), Levitt (32) and Chou and Fasman (33) on all six proteins correctly predict the first of the two α-helices terminating at the conserved glycine. The following β-turn is either predicted as a β-turn or as β-sheet. The second helix can only be found in the predictions for λ-cro,

Table 2. Pairwise comparison of identities in the DNA-binding regions of the six repressors shown in Fig. 4.

	λcro	λcI	P22cI	lacI	galR	deoR
λcro	-					
λcI	22	-				
P22cI	35	30	-			
lacI	13	22	25	-		
galR	17	22	25	61	-	
deoR	22	17	30	39	48	-

All comparisons are for the 23-residue region shown in Fig. 4, except for P22cI of which only a 20-residue region has been published

while the predictions for the other repressors indicate that this region should be in β-sheet conformation. As this is known from x-ray studies not to be the case for the λ-repressor and CAP (24,26) some factors not taken into account by the prediction methods must be determining the secondary structure in this region.

The strong homology and similar secondary structure predictions suggest that residues number 22-44 in DeoR are involved in binding to the DNA helix.

Regulation of the deo operon by DeoR

The DeoR repressor has been shown to control two promoter regions (P1 and P2), located 600 bp apart, in front of the deo operon (9-11) (Fig. 1). The operator sites have been defined by construction of gene fusions between the P1 and P2 regions and lacZ (10), and by sequence homology (9). The targets most likely consist of a 16 bp palindrome which encompasses the Pribnow box sequence of each of the two promoters. Out of the 16 bp only one position differs between the targets (Fig. 5). Interestingly, a region located just downstream from the start of deoR (from bp 38 to 53) shows a striking homology with these two targets for DeoR (Fig. 5). Whether or not this region is functionally important in deoR gene expression remains to be determined.

One of the most perplexing features of the deo operon regulation is the strong increase in expression of the deo enzymes when going from a single regulatory mutant (deoR$^-$ or cytR$^-$ strains) to a double regulatory

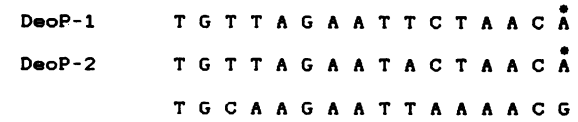

Figure 5. Comparison of the putative DeoR target in the deoP1 and P2 promoter regions with a sequence of deoR located from bp 38 to 53. The P1 and P2 transcriptional start sites are indicated by asterisks.

mutant (deoR$^-$ and cytR$^-$ strain) (8). The level of deo enzymes produced in a double regulatory mutant is thus found to be 4- to 5-fold higher than the sum of the levels observed in each of the single regulatory mutants. How can this regulatory feature be explained? We strongly suggest that the regulation of both P1 and P2 expression by DeoR is responsible for this property of deo regulation, since 1) the deoP2 promoter seems to be much stronger than deoP1, 2) strong repression of P2 by DeoR is observed in fusions between the P1-P2 region and lacZ. However, our studies of gene fusions have also shown that in absence of the P1 region DeoR regulation of the P2 promoter is weak. Thus these results indicate that full repression of deoP2 by DeoR requires not only the deoP2-DeoR target but also the P1 target. Recently, it has been shown that repression of gal requires two gal repressor binding sites separated by about 110 bp (34-35) and also two targets for AraC are required for maximal repression of the araBAD promoter (36). How interaction takes place between tandemly repeated operators separated by more than hundred base pairs is unknown. Experiments to define the features of deoR repressor binding and its involvement in deo regulation are currently in progress.

ACKNOWLEDGEMENTS

We thank Ulla Pedersen for carefully typing the manuscript, Marianne Hald for technical assistance and Roger S. Buxton and Karin Hammer for supplying the Sø3135 strain. This research was supported by grants from The Carlsberg Foundation and The Danish Natural Science Research Council to P. Valentin-Hansen.

*To whom reprint requests should be sent

REFERENCES

1. Hammer-Jespersen, K. (1982) in Metabolism of Nucleotides, Nucleosides and Nucleobases in Microorganisms, Munch-Petersen, A. Ed., Academic Press, London, pp. 203-258.

2. Munch-Petersen, A. and Mygind, B. (1982) in Metabolism of Nucleotides, Nucleosides and Nucleobases in Microorganisms, Munch-Petersen, A. Ed., Academic Press, London, pp. 259-305.
3. Breitman, T.R. and Bradford, R.M. (1967) Biochim. Biophys. Acta (Amst.) 138, 217.
4. Svenningsen, B.A. (1975) Molec. Gen. Genet. 137, 289-304.
5. Valentin-Hansen, P., Svenningsen, B.A., Munch-Petersen, A. and Hammer-Jespersen, K. (1978) Molec. Gen. Genet. 159, 191-202.
6. Munch-Petersen, A. and Mygind, B. (1976) J. Cell. Physiol. 89, 551-560.
7. Krieger-Brauer, H.J. and Braun, V. (1980) Arch. Microbiol. 124, 233-242.
8. Hammer-Jespersen, K. and Munch-Petersen, A. (1975) Molec. Gen. Genet. 137, 327-335.
9. Valentin-Hansen, P., Aiba, H. and Schümperli, D. (1982) EMBO J. 1, 317-322.
10. Valentin-Hansen, P. (1985) in Gene Manipulation and Expression, Glass, R.E. and Spizek, J. Eds. Croom Helm., in press.
11. Short, S.A. and Singer, J.T. (1984) Gene 31, 205-211.
12. Norrander, J., Kempe, T. and Messing, J. (1983) Gene 27, 101-106.
13. Shine, J. and Dalgarno, L. (1975) Proc. Natl. Acad. Sci. USA 72, 4734-4738.
14. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-354.
15. Siebenlist, V., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
16. Hawley, D.K. and McClure, W.R. (1983) Nucleic Acids Res. 11, 2237-2255.
17. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1981) Ann. Rev. Microbiol. 35, 365-403.
18. Gilson, E. Clement, J.M., Brutlag, D. and Hoffnung, M. (1984) EMBO J. 3, 1417-1421.
19. Valentin-Hansen, P., Hammer-Jespersen, K., Boëtius, F. and Svendsen, I. (1984) EMBO J. 3, 179-183.
20. Stern, M.J., Ames, G.F., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) Cell 37, 1015-1026.
21. Grantham, R., Gautier, C., Gouy, M., Mercier, R and Pave, A. (1980) Nucleic Acids Res. 8, r49-r62.
22. Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Natl. Acad. Sci. USA 76, 1697-1701.
23. Farabough, P.J. (1978) Nature 274, 765-769.
24. McKay, D. and Steitz, T. (1981) Nature 290, 744-749.
25. Anderson, W., Ohlendorf, D., Takeda, Y. and Matthews, B. (1981) Nature 290, 754-758.
26. Pabo, C. and Lewis, M. (1982) Nature 298, 443-447.
27. Weber, I.T. and Steitz, T.A. (1984) Proc. Natl. Acad. Sci. USA 81, 3973-3977.
28. Sauer, R.T., Yocum, R.R., Doolittle, R.T., Lewis, M. and Pabo, C.O. (1982) Nature 298, 447-451.
29. Gicquel-Sanzey, B. and Cossart, P. (1982) EMBO J. 1, 591-594.
30. Pabo, C.O. (1984) Ann. Rev. Biochem. 53, 293-321.
31. Garnier, J., Osguthope, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
32. Levitt, M. (1978) Biochemistry 17, 4277-4285.
33. Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 222-245.
34. Irani, M.H. Orosz, L. and Adhya, S. (1983) Cell 32, 783-788.
35. Majumdar, A. and Adhya, S. (1984) Proc. Natl. Acad. Sci. USA 81, 6100-6104.
36. Dunn, T.M., Hahn, S., Ogden, S. and Schleif, R.F. (1984) Proc. Natl. Acad. Sci. USA 81, 5017-5020.