Volume 5 Number 10 October 1978 Nucleic Acids Research

A computer aided oligonucleotide analysis provides a model sequence for RNA polymerase - promoter recognition in E.coli

Günther E.F.Scherer, Malcolm D.Walkinshaw* and Struther Arnott

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA and
*Max Planck Institut für Experimentelle Medizin, Hermann-Rein-Strasse 3, D-3400 Göttingen, GFR

Received 3 August 1978

ABSTRACT

A novel computer procedure has been used to search for homology among 17 known procaryotic promoter sequences. A model sequence, acc·t·gttGTTGAcATTTtt····ttggcGGTTATATTg···cCAT is
t a a ATA a g
compatible with the properties of *all* known promoter and operator mutations, predicts base positions for the initiation of RNA synthesis coinciding with those determined experimentally, is compatible with current models for the regulation of transcription, suggests that RNA polymerase could recognize the DNA double helix firstly in the *B* conformation then in the *A*.

INTRODUCTION

To begin RNA transcription, the RNA polymerase of *Escherichia coli* recognizes and binds to the promoter region of a gene. Base sequences for 17 such regions are now known[1-28] (Fig. 1). Since they all interact with the same enzyme,[29] an intrinsic similarity should exist among them. Indeed, Pribnow[26] has suggested that a sequence like TATPuATPu should occur near the site of initiation of RNA synthesis, and Gilbert[33] has noted that the duodecanucleotide TGTTGACAATTT is observed in *many* but *not all* promoters at a position more distant from the initiation site. However, it is evident that there is rarely an absolute requirement for a particular base in a particular position with respect to the common origin. Rather, it appears that there is a *preference* for a particular base at specific sites. In these circumstances, there should exist an ideal promoter sequence to which all promoters approximate. The requirement of close approximation may be greater in some blocks of this sequence than others. To identify such a sequence intuitively by inspection is not easy. What is needed is a more objective procedure which quantitates the rela-

tive importance of a particular base at a given position. The method described below provides this. The model sequence provided by its application to 17 promoter sequences not only contains blocks reminiscent of those mentioned above but also suggests that the identity of bases at 19 other positions may be of some importance.

METHOD

Each DNA sequence was lined up and the base positions numbered as in Fig. 1.

Blocks of 3 or 4 bases (corresponding to the trimer or tetramer analyses discussed below) were stepped in 1 base increments along each sequence. All possible block sequences were tried. Wherever the bases in both block and strand matched, the block composition was noted. The resulting list was culled leaving only blocks which occurred in more than one strand and/or were $\leq|x|$ base positions distant. (The "block deviation", x, was 2 in the trimer analysis and 3 in the tetramer analysis.)

All combinations of identical blocks occurring in two or more promoter sequences within the allowed block deviations were taken two at a time and scores of 1, 1/3, 1/5, or 1/7 assigned to each base depending on whether the relative displacement of the blocks was 0, 1, 2, or 3. For each base position, the base scores accumulated for all block pairings were summed. The total score, Σ, was graphed. (A simple example of the procedure is illustrated in Table 1.)

In the application to the promoter sequences, the effect of varying the block size was tested by conducting both trimer and tetramer searches. Since the same model sequence with the same relative peak heights emerged from both analyses, further discussion will be confined to the results of the trimer analysis shown in Fig. 2a.

The level above which values of Σ should be considered significant was estimated empirically by applying the procedure to 17 sequences of "nonsense" DNA. Each of these sequences had the same number of bases as a promoter counterpart. The results of this analysis (shown in Fig. 2b) suggest that values of $\Sigma \geq 18$ are likely to be significant. (Where $n < 17$ strands contributed

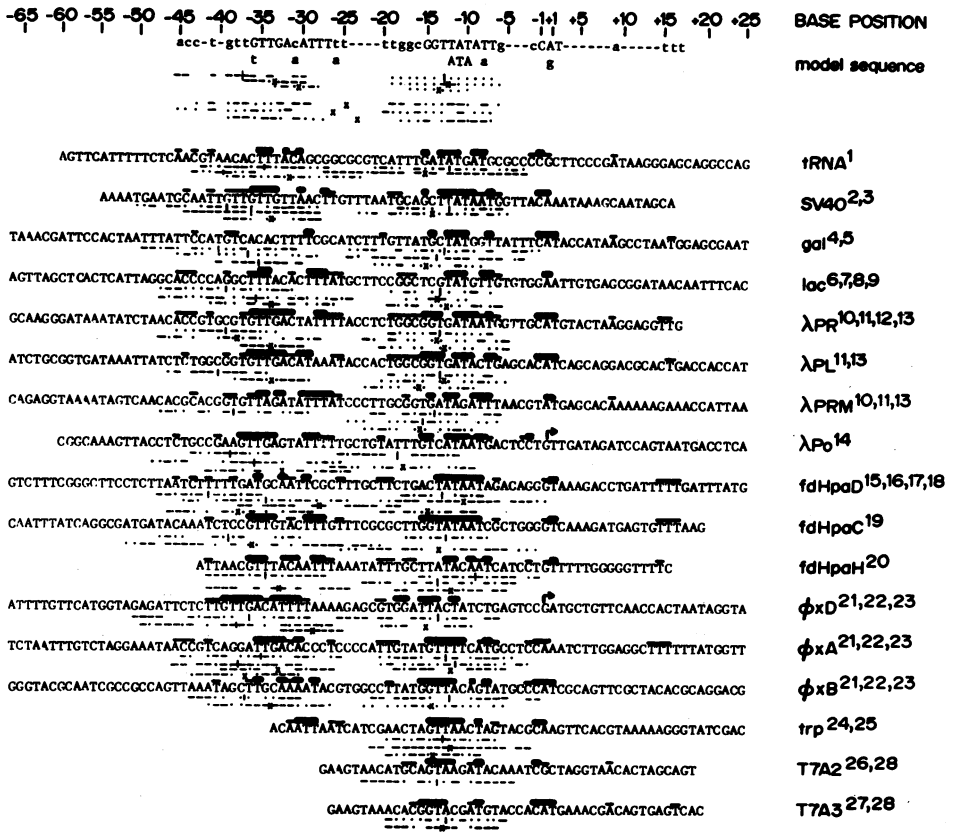

Figure 1 - Comparison of the model sequence with the individual promoter sequences. The base position numbers are shown on top; as a reference position the 5' end of the transcribed RNA was chosen and designated +1. Sequences which are not transcribed are given negative position numbers. The fd Hpa C(G3) promoter starting position (determined by oligonucleotide priming) is heterogenous. Our analysis indicated a best fit as shown. As ϕXB promoter we took the sequence which is discussed in ref. 21 to be most likely the true promoter. Starting base is nucleotide number 4888 in the ϕX sequence. All references are given for sequence and starting base(s) in the Figure.

An exact base to base correspondence between the model sequence and promoters is indicated by thick (capital letters) and thin bars (small letters). For the ϕXD and λp_O promoter the best fit instead of base correspondence for the reference base is shown (see Table 2). Below the sequences, centers of twofold symmetry are indicated by a X, "pseudopalindromes" are indicated by a |. Exact base/base correspondence in these symmetry properties are underlined, purine/pyrimidine correspondence is indicated by dots.

TABLE 1

Base Position	Sequences 1 2 3 4 5 6 7 8 9	Σ values 1 2 3 4 5 6 7 8 9	
	A C C A T G A A C	2 0 0 $\frac{1}{3}$ $\frac{1}{3}$ 0 6 6 0	Σ(A)
	G T G C A T T T G	0 0 0 0 0 0 0 0 0	Σ(G)
	T C A G C C A A C	0 2 $2\frac{1}{3}$ $\frac{1}{3}$ 0 0 0 0 6	Σ(C)
	A C C T C T A A C	0 0 0 0 $\frac{1}{3}$ $\frac{1}{3}$ 0 0 0	Σ(T)

Table 1 - Determination of Σ values in a simple, arbitrary example. All contributing trimers are underlined. Note the increase in Σ values for the bases in the AAC trimer due to the formation of three pairs of trimers.

to the values of Σ at a particular site, the threshold was taken to be 18n/17.) Further, since almost all known promoter mutants occur in positions where Σ ≥ 30 and all operator mutants are without effect on positions with high values of Σ, we have chosen to ascribe a higher level of significance to positions with Σ ≥ 30.

Since the promoter sequences investigated have a relatively high AT content, we repeated the experiment with random sequences constrained to have a 60% AT content in the region from -45 to +1. This suggested that for A or T, a significant Σ should be ≥ 20 and for G or C, Σ should be ≥ 12. For the discussion below, we have retained Σ = 18 as the threshold for significance although we recognize that in some circumstances this value may be slightly liberal with respect to A and T and rather conservative with respect to G and C.

RESULTS AND DISCUSSION

The model sequence. Our analysis (Fig. 2a) suggests that the sequence

```
acc-t-gttGTTGAcATTTtt----ttggcGGTTATATTg---cCAT------a-----ttt
         t    a    a              ATA a       g
```

contains bases which have functional importance when they are in these relative positions within promoters. The lower case letters

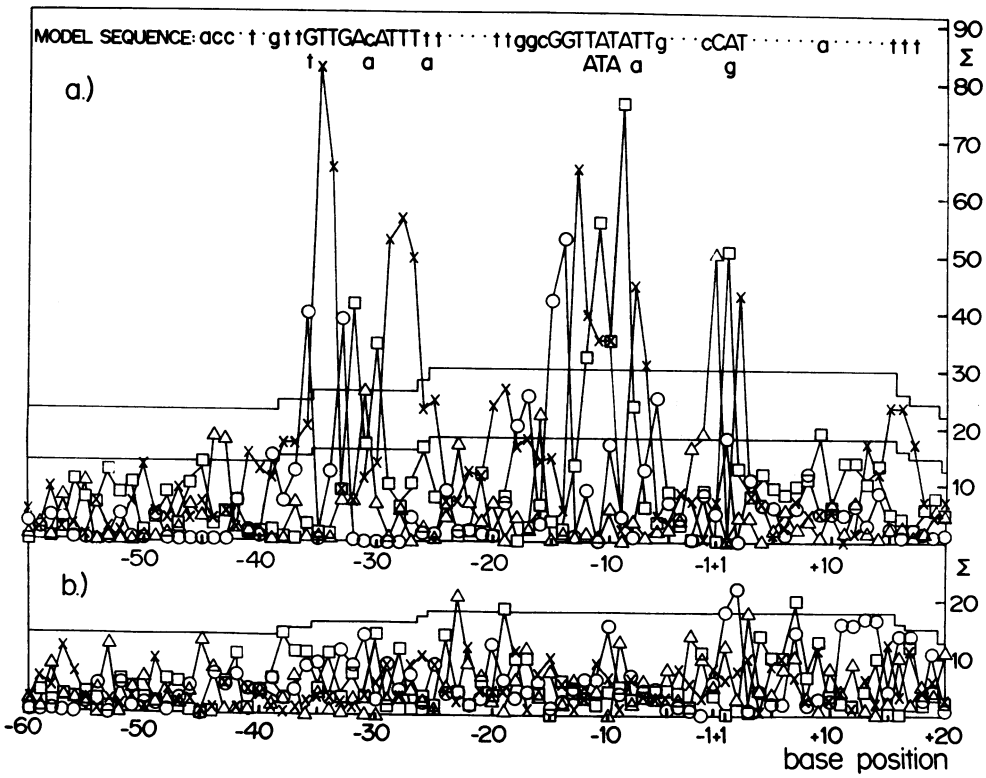

Figure 2 - Trimer analysis (with block deviation ±2) for 17 promoter sequences and 17 random DNA sequences of corresponding length. a) The promoters were lined up as in Fig. 1. The graph shows the result of a trimer analysis on 17 promoter sequences. The model sequence is shown above. b) Result of the same type of analysis of random DNA sequences. Σ(A):□; Σ(G):O; Σ(C):Δ; Σ(T):X .

correspond to peaks with $18 \leq \Sigma \leq 30$ and the upper case letters to peaks with $\Sigma > 30$.

There are three blocks containing bases with $\Sigma > 30$ which appear to have high functional significance. One with the sequence GTTATATT(g) has the Pribnow[26] sequence at one end. For
ATA a
3 and possibly 4 of the positions in this sequence there appears to be a requirement for an A:T base-pair in the DNA duplex rather than for a particular base in one strand. Another, around position -30, is similar to Gilbert's[33] duodecanucleotide. Sequenced promoter mutants,[7,30-33] binding studies[7,11,12,15,16,19,24,26,27

and digestion experiments with different DNAases[11,34-36] have pointed to the importance of this region and similar sequences have been proposed by a number of other workers.[14,20] Both of these sequences have AT tails pointing in the direction of RNA synthesis. This introduces an interesting polarity and suggests that local melting properties may have a role. In this connection, it is worth pointing out that our model sequence is flanked by two regions of high AT composition: positions -45 to -56 contain 68% AT; positions -21 to +13 have 63% AT.

The third block of high significance is a triplet CAT at the site of RNA initiation. The importance of the C at position -1 has been evident from oligonucleotide priming experiments.[23,36]

Symmetry properties. The significance of local symmetry elements in promoter sequences has been discussed frequently.[1,10,11,13,15,16,19,26,37] In our model sequence, repetitiveness and local dyadic symmetry are not so pronounced that we would regard them as strong requirements for promoter binding although undoubtedly they would enhance this if some of the $\alpha_2\beta\beta'\sigma$ subunits of the holoenzyme are related by quasi-dyads.

The effect of mutations. If indeed the bases in our model sequence are favorable for efficient promoter function, there should be a predictable correlation between promoter mutations and their effects. In more efficient mutants, the base change(s) would have produced a closer agreement with the model sequence. In less efficient mutants, the reverse would be the case. Six out of seven sequenced promoter mutants clearly conform with this prediction (see Fig. 3).

In the two lac mutations with enhanced promoter efficiency (lac UV5 and lac p^S), a G in the wild type sequence is replaced by a highly preferred A at -9. In four other mutants with diminished promoter efficiency (lac L305, lac L241, λprm116, λsex1) the base changes reduce similarity to our model sequence. In the one dubious case (as far as our hypothesis is concerned) a C→A mutation at position -16 *increases* promoter efficiency under CRP^- conditions but *reduces* it under CRP^+ conditions. Perhaps in the absence of the CRP, a lowering of the DNA melting point becomes more important than base recognition.

For none of the operator mutants has the influence on

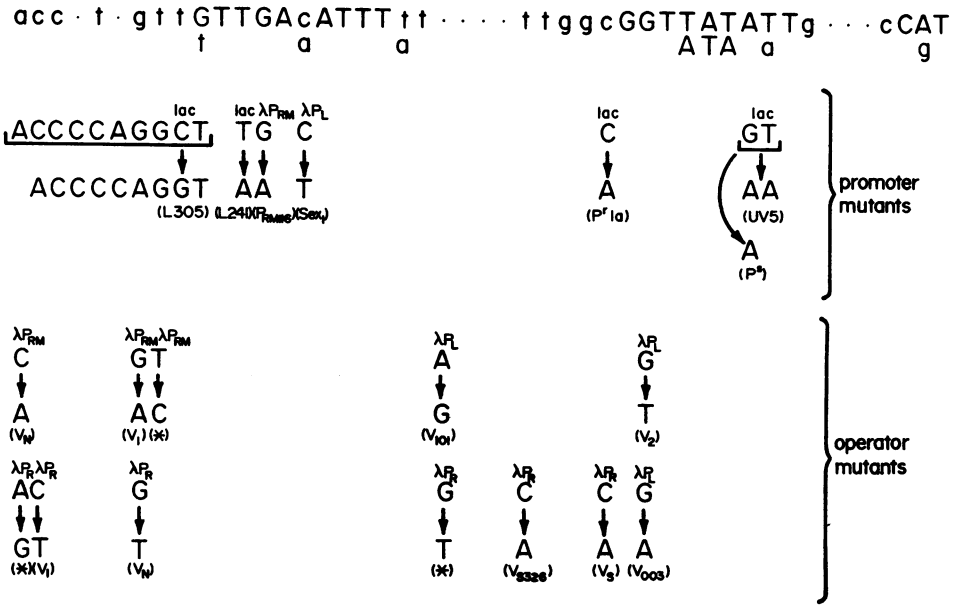

Figure 3 - Comparison of the model sequence with all known mutations in promoter regions. The model sequence is shown on top. All known promoter and operator mutations are listed below, using the 5' end of the respective RNA's as a reference point as in Fig. 1 (see references 7 and 33 for mutants in the lac promoter and 30-32 for the λ promoters). Wild type promoters are designated above the wild type base, the mutant base is indicated by an arrow. The designation of the mutant is given in brackets below the mutant base ((*) = no designation).

promoter function been characterized quantitatively. However, the viability of the twelve whose sequences are known is understandable in terms of our model. In five of the twelve, the mutations neither enhance nor reduce the similarity to the model sequence. In another five, the mutations do not fall within what we consider to be significant regions. In two λP_L mutations, G at position -14 is replaced by A and thus extends the 8 A,T base-pair track (at position -13 through -7) by one base-pair. Once again it is possible that easier local melting would be a factor competing with precise base recognition.

Initiation. The requirement for a purine at the site of initiation of RNA synthesis is very strong. However, heterogeneity of the 5' ends of RNAs has been found.[8,9,16,18,19,22,23] Both these observations are consistent with our model sequence.

We have compared (Table 2) each promoter sequence with the model sequence, shifting the Pu in the CAT signal from base positions designated -5 through +6 in Fig. 1. In nearly every case, the best match is obtained when +1 coincides with A, Exceptions like the ϕXD promoter, which gives the best fit at -3 but initiates RNA synthesis at the first upfield purine, highlight the requirement for this kind of base.

In a number of cases, there are a number of sites clustered around +1 which provide both the required purine and a good fit with the model sequence. These multiple opportunities for the formation of a successful initiation complex may result in variable 5' ends for RNA.

Promotion and repression. That our model sequence suggests that the promoter region is longer than previously considered is quite compatible with a number of experiments on the regulation of transcription.[7,10,11,33,41] For example,[40] cutting the t-RNA promoter at position -33/-34 decreases transcription markedly compared with cutting it at position -43/-44. In addition, the properties of the deletion mutant lac L305 indicate that elements beyond position -38 are necessary for efficient promoter function.[7,39]

Positive regulation of the transcription of gal and lac m-RNA by the CRP protein has been demonstrated.[4,7] This protein probably covers positions -75 to -50 or -48. Our suggestion that the interaction site for RNA polymerase might extend as far as position -45 would not be incompatible with the proposal[33] that the CRP protein could help "lock" the polymerase to the interaction site. The other suggestion[7] that CRP might act by inducing an extensive conformational change in DNA (and thereby alter physical properties, such as melting) is not precluded.

Indeed, the requirement of an activating protein by the gal and lac promoter is understandable when one considers their considerable deviation from our ideal sequence. In gal, the signal around position -30 is almost totally missing and the "Pribnow box" contains 2 Gs. Although the lac promoter is somewhat more similar to the model sequence, it has a lower (50%) AT content, whereas other promoters have 60-70% AT.

TABLE 2

	-5			-4			-3			-2			-1			+1			+2			+3			+4			+5			+6		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
tRNA	9	6	G	5	8	C	6	3	C	5	3	C	7	5	C	11	7	G	5	7	C	7	6	T	2	8	T	3	4	T	3	4	C
SV40	7	5	G	7	7	T	9	9	T	7	5	A	5	8	C	15	10	A	11	9	A	8	9	A	11	9	T	8	5	A	6	3	G
gal	10	6	A	11	7	T	8	9	T	8	8	T	4	7	C	9	6	A	4	8	T	4	4	A	6	6	C	9	4	C	14	8	A
lac	4	2	T	4	4	G	6	8	T	8	8	G	9	8	G	11	12	A	11	8	A	7	7	T	5	5	T	6	10	G	7	10	T
λP_R	4	4	G	3	4	T	7	7	T	7	7	G	6	12	C	17	17	A	11	10	T	8	6	G	10	6	T	7	5	A	4	3	C
λP_L	5	3	A	1	5	G	7	5	C	10	5	A	5	8	C	17	9	A	3	11	T	6	5	C	9	8	A	5	7	G	4	3	C
λP_{RM}	4	5	A	10	5	A	6	6	C	13	8	G	10	11	T	15	9	A	8	8	T	6	7	G	7	7	A	8	6	G	5	7	C
λP_O	4	5	C	8	9	T	6	7	C	7	8	C	14	9	T	11	8	G	8	7	T	9	10	T	7	9	G	8	4	A	4	5	T
fdHpaD	6	12	A	7	11	C	8	8	A	6	12	G	12	11	G	10	13	G	4	10	T	8	9	A	6	8	A	8	3	A	7	9	G
fdHpaC	8	6	C	7	9	T	5	11	G	3	10	G	10	10	G	16	8	G	7	7	T	7	10	C	9	11	A	6	7	A	5	3	A
fdHpaH	9	3	A	5	5	T	4	9	C	9	6	C	7	8	T	12	11	G	12	7	T	7	5	T	7	7	T	5	4	T	4	5	T
ϕXD	7	8	A	10	9	G	15	5	T	9	8	C	9	5	C	6	13	G	6	8	A	5	4	T	5	8	G	9	6	C	8	5	T
ϕXA	7	3	C	10	3	C	6	5	T	6	9	C	6	9	C	13	15	A	9	9	A	7	8	A	4	8	T	4	7	C	4	4	T
ϕXB	7	7	T	2	9	G	2	7	C	7	6	C	10	4	C	14	4	A	7	9	T	4	9	C	8	6	G	3	5	C	7	4	A
trp	5	4	T	4	2	A	4	5	C	5	7	G	9	4	C	11	4	A	7	4	A	7	3	G	6	6	T	4	8	T	7	3	C
T7A2	5	1	A	2	6	A	5	3	A	4	1	T	4	2	C	7	5	G	4	5	C	4	4	T	4	5	A	6	5	G	9	5	G
T7A3	3	2	A	1	4	C	3	3	C	3	3	A	4	3	C	9	4	A	3	3	T	3	5	G	4	4	A	5	2	A	7	2	A

Table 2 - Predictability of the starting base of the 5' end of the RNA by base:base comparison of the model sequence and individual promoter sequences. The model sequence was lined up against the individual promoter strands with the adenosine in the CAT signal at each of the base positions -5 through +6 (reference position +1 as shown in Fig. 1). At each base position the number of matching upper and lower case letters are given in columns a and b, respectively. Column c indicates the base with which transcription would start. Experimentally determined starting positions are underlined. Heterogenous starting positions are shown by dotted lines. (For references, see Fig. 1.)

In addition, it is known[41] that the gal repressor is inefficient as compared with the lac[42] and λ repressors.[43,44] Although it has been suggested[41] that the gal repressor acts by preventing stimulation of transcription by the CRP protein, our model sequence suggests another explanation. A study of the gal operator sequence[4] showed a region between positions -77 and -46 with a diad axis enclosing O^c mutants. If the gal repressor covers this region and extends even slightly beyond it, steric hindrance may preclude transcription starting at the best sites (positions -5 without CRP and +1 with CRP). However, a start at position +6 is a possible (less efficient) alternative (see Table 2). Since it has not yet been determined experimentally whether this is indeed the starting position for the gal promoter under repression conditions, this is a prediction which can still be tested.

Again, the length of our model sequence suggests that one of the two mechanisms proposed[10,11] for the regulatory switch of RNA synthesis between the adjacent promoters λp_R and λp_{RM} is more likely to be correct. It has been determined experimentally[11] that RNA polymerase binds preferentially to λp_R (which is also what our analysis presented in Table 2 would predict). If the promoter region is about 45 bases long, then the polymerase would act as an autogenous repressor because both promoters overlap by 8 bases (Fig. 4d). The binding of the λ repressor to the stronger operator OR1 would both remove this autogenous inhibition and inhibit transcription at the λp_R promoter (Fig. 4e). The combined effect would be *apparent* positive regulation.

The DNA conformations involved. The main conformational variants of DNA are *B* and *A* which have similar numbers of base-pairs per helix pitch (10 and 11, respectively) but quite different pitch lengths (33.7Å and 28.2Å, respectively). Direct recognition of either conformation by proteins has been shown to be possible by model-building.[45-47] The involvement of the RNA-like *A*-form in transcription has been postulated before.[48-52] The available experimental evidence suggests that both the *B* and *A* conformations might be involved, probably sequentially.

The *B* conformation is favored in relatively hydrophilic environments and is likely to be the form which has to be recognized initially by any interacting protein.[49] The *A* conformation

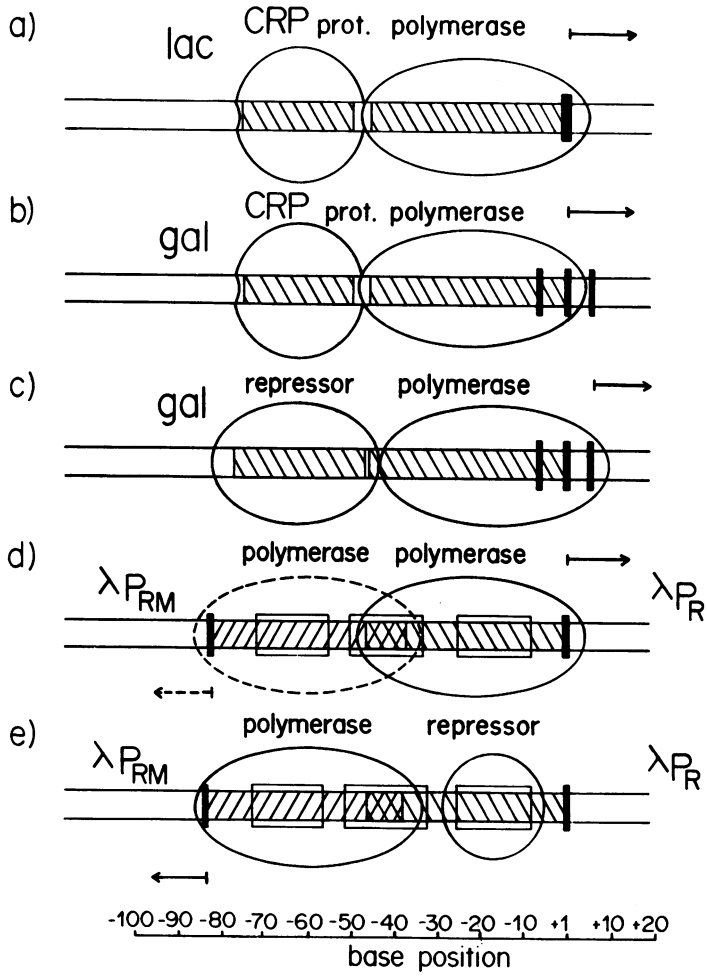

Figure 4 - Postulated steric arrangement of RNA polymerase and regulatory proteins at some promoters. Binding sites on DNA strands are drawn as crosshatched boxes, the binding sites of the λ repressors as boxes slightly bigger than the DNA. Experimentally determined (and hypothetical) initiating bases are indicated as vertical bars on the DNA strand. Arrows indicate the start and direction of transcription. The shapes of the proteins are schematic. Proteins and promoters are designated in the figure. Dashed lines indicate steric hindrance. See text for references of original literature.

is favored by relatively hydrophobic environments and G,C-rich base sequences. It is interesting, therefore, that the relatively G,C-rich lac promoter in the presence of glycerol can be tran-

scribed *in vitro* relatively more efficiently than other promoters.[39] The t-RNA promoter is salt sensitive and *in vitro* must also be activated by materials like glycerol.[36]

The *B→A* transition unwinds DNA 9% and there is evidence that RNA polymerase partially unwinds DNA during complex formation with fdDNA.[53] Unwinding is indispensible also for tight binding and profoundly influences the binding constant for the SV40 promoter.[54]

The *B→A* transition has two further effects on DNA: the length is reduced by 16% and A,T-rich regions become unstable. In these circumstances, the involvement of bases over the long length from position -45 to +2 (158Å in *B*-DNA, but only 120Å in *A*-DNA) becomes less surprising. One could envision, for example, that the first stage of RNA polymerase binding would involve bases in only two of the significant regions. If this induced a *B→A* transition, then the important block around +1 could also be accommodated in a binding site of essentially the same length as before. At this stage, the duplex structure in the A,T-rich blocks between positions -36 and -24 and between positions -14 and -6 would become unstable and the necessary local melting of the DNA could begin.

CONCLUSION

Our method of sequence analysis is novel in that it does not seek for complete homology but instead identifies sequences which are approximately homologous and therefore have analogous properties. It shows that the sequences in certain regions of promoters are more important than others and defines an "ideal" sequence for these blocks. The extent to which individual promoters and their mutants approach this ideal sequence can be used to predict their relative efficiency in transcription.

Addendum. During the course of the revision of this paper, the sequences of another three procaryotic promoters were published.[55-59] One promoter originates from the insertion mutation c17 in the λ genome. RNA polymerase starts transcription in this promoter with an unusual base, cytosine. The best agreement of our model sequence with this promoter is achieved if

exactly this cytosine is used as position +1. Then thirteen bases of this promoter agree with the bases written in upper case letters in our model sequence and 6 bases agree with bases written in lower case letters. This is comparable to the values found for the other promoters (Table 2). Such good agreement can not be found with the wild-type sequence.

The trp promoter sequence of *E. coli*[56] has been extended up to position -118, and the very similar trp promoter of *S. typhimurium* has been sequenced. The extended promoter regions contain the sequences TGTTGACATTATT (in *E. coli*) and TGTTGACAATTAA (in *S. typhimurium*), both in excellent agreement with our model sequence. Moreover, functional studies show that a region covering at least positions +1 to -39 is needed for promoter function,[58] and all of the six O^c mutations falling within the promoter region again do not negatively affect the similarity of the trp promoter with the model sequence.[59]

ACKNOWLEDGEMENTS

We thank Professor D. James Morré for encouragement. This work was supported by the Max-Kade Foundation (through a fellowship to G.S.) and by NIH (through a grant no. GM 17371 to S.A.).

REFERENCES

1 Sekiya, T., Gait, M.J., Noris, K., Ramamoorthy, B. and Khorana, H.G. (1976) J. Biol. Chem. 251,4481-4489
2 Zain, B.S., Weissman, S.M., Dhar, R. and Pan, J. (1974) Nucleic Acids Res. 1,577-594
3 Dhar, R., Weissman, S.M., Zain, B.S., Pan, J. and Lewis Jr., A.M. (1974) Nucleic Acids Res. 1,595-613
4 Musso, R., Di Lauro, R., Rosenberg, M. and de Crombrugghe, B. (1977) Proc. Natl. Acad. Sci. USA 74,106-110
5 Musso, R.E., de Crombrugghe, B., Pastan, I., Sklar, J., Yot, P. and Weissmann, S. (1974) Proc. Natl. Acad. Sci. USA 71, 4940-4944
6 Gilbert, W. and Maxam, A. (1973) Proc. Natl. Acad. Sci. USA 70,3581-3584
7 Dickson, R.C., Abelson, J., Barnes, W.M. and Reznikow, W.S. (1975) Science 187,27-35
8 Maizels, N.M. (1973) Proc. Natl. Acad. Sci. USA 70,3585-3589
9 Majors, J. (1975) Proc. Natl. Acad. Sci. USA 72,4394-4398
10 Walz, A., Pirrotta, V. and Ineichen, K. (1976) Nature 262, 665-669
11 Ptashne, M., Backmann, K., Humayun, M.Z., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, R.T. (1976) Science 194, 156-161
12 Walz, A. and Pirrotta, V. (1975) Nature 254,118-121

13 Blattner, F.R. and Dahlberg, J.E. (1972) Nature New Biol. 237,227-232
14 Scherer, G., Hobom, G. and Kössel, H. (1977) Nature 265,117-121
15 Schaller, H., Gray, C. and Herrmann, K. (1975) Proc. Natl. Acad. Sci. USA 72,737-741
16 Sugimoto, K., Okamoto, T., Sugisaki, H. and Takanami, M. (1975) Nature 253,410-414
17 Sugimoto, K., Sugisaki, H., Okamoto, T., and Takanami, M. (1975) Nucleic Acids Res. 2,2091-2100
18 Heyden, B., Nüsslein, C. and Schaller, H. (1975) Eur. J. Biochem. 55,147-155
19 Takanami, M., Sugimoto, K., Sugisaki, H. and Okamoto, T. (1976) Nature 260,297-302
20 Seeburg, P.H., Nüsslein, C. and Schaller, H. (1977) Eur. J. Biochem. 74,107-113
21 Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison III, C.A., Slocombe, P.M. and Smith, M. (1977) Nature 265,687-695
22 Smith, L.H. and Sinsheimer, R.L. (1976) J. Mol. Biol. 103, 711-735
23 Axelrod, N. (1976) J. Mol. Biol. 108,753-770
24 Bennett, G.N., Schweingruber, M.E., Brown, K.D., Squires, C. and Yanowsky, C. (1976) Proc. Natl. Acad. Sci. USA 73,2351-2355
25 Squires, C., Lee, F., Bertrand, K., Squires, C.L., Bronson, M.J. and Yanowski, C. (1976) J. Mol. Biol. 103,351-381
26 Pribnow, D. (1975) Proc. Natl. Acad. Sci. USA 72,784-788
27 Pribnow, D. (1975) J. Mol. Biol. 99,419-443
28 Kramer, R.A., Rosenberg, M. and Steitz, J.A. (1974) J. Mol. Biol. 89,767-776
29 Bautz, E.K. (1973) FEBS Letters 36,123-129
30 Meyer, B.J., Kleid, D.G. and Ptashne, M. (1975) Proc. Natl. Acad. Sci. USA 72,4785-4789
31 Kleid, D., Humayun, Z., Jeffrey, A. and Ptashne, M. (1976) Proc. Natl. Acad. Sci. USA 73,293-297
32 Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A. and Maurer, R. (1975) Cell 5,109-113
33 Gilbert, W. (1976) in RNA Polymerase (Losick, R. and Chamberlin, M., eds.) pp. 193-205. Cold Spring Harbor Laboratory, New York
34 Maurer, R., Maniatis, R. and Ptashne, M. (1974) Nature 249, 221-223
35 Allet, B., Roberts, R.J., Gesteland, R.F. and Solem, R. (1974) Nature 249,217-221
36 Küpper, H., Contreras, R., Khorana, H.G. and Landy, A. (1976) in RNA Polymerase (Losick, R. and Chamberlin, M., eds.) pp. 473-484. Cold Spring Harbor Laboratory, New York
37 Dykes, G., Bambara, R., Marians, K. and Wu, R. (1974) Nucleic Acids Res. 2,327-344
38 Arditti, R., Grodzicker, T. and Beckwith, J. (1973) J. Bact. 114,652-655
39 Reznikoff, W.A. (1976) in RNA Polymerase (Losick, R. and Chamberlin, M., eds.) pp. 441-454. Cold Spring Harbor Laboratory, New York
40 Sekiya, T., Takeya, T., Contreras, R., Küpper, H., Khorana, H.G. and Landy, A. (1976) in RNA Polymerase (Losick, R. and

Chamberlin, M., eds.) pp. 455-472. Cold Spring Harbor Laboratory, New York
41 Nakanishi, S., Adhya, S., Gottesman, M. and Pastan, I. (1973) J. Biol. Chem. 248,5937-5942
42 de Crombrugghe, B., Chen, B., Anderson, W., Nissley, P., Gottesman, M. and Pastan, I. (1971) Nature New Biol. 231, 139-142
43 Steinberg, R.A. and Ptashne, M. (1971) Nature New Biol. 230, 76-80
44 Wu, A.M., Gosh, S., Echols, H. and Spiegelman, W.G. (1972) J. Mol. Biol. 67,407-421
45 Blake, C.C.F. and Oatley, S.J. (1977) Nature 268,115-120
46 Carter Jr., C.W. and Kraut, J. (1974) Proc. Natl. Acad. Sci. USA 71,283-287
47 Church, G.M., Sussman, J.L. and Kim, S.-H. (1977) Proc. Natl. Acad. Sci. USA 74,1458-1462
48 Arnott, S., Fuller, W., Hodgson, A. and Prutton, I. (1968) Nature 220,561-564
49 Arnott, S. (1976) in Organisation and Expression of Chromosomes (Allfrey, V.G., Bautz, E.K.F., McCarthy, B.J., Schimke, R.T. and Tissieres, A., eds.) pp. 209-222. Life Science Research Reports, Vol. 4, Berlin: Dahlem Konferenzen
50 Arnott, S., Bond, P.J., Selsing, E. and Smith, P.J.C. (1976) Nucleic Acids Res. 3,2459-2469
51 Bram, S. and Tougard, P. (1972) Nature New Biol. 239,128-131
52 Florentiev, V.L. and Ivanov, V.I. (1970) Nature 228,519-522
53 Wang, J.C., Jacobsen, J.H., Saucier, J.-M. (1977) Nucleic Acids Res. 4,1225-1241
54 Warner, C.K. and Schaller, H. (1977) FEBS Letters 74,215-219
55 Rosenberg, M., Court, D., Shimatake, H., Brady, C. and Wulff, D.L. (1978) Nature 272,414-421
56 Bennett, G.N., Schweingruber, M.E., Brown, K.D., Squires, C. and Yanowsky, C. (1978) J. Mol. Biol. 121,113-137
57 Bennett, G.N., Brown, K.D. and Yanowsky, C. (1978) J. Mol. Biol. 121,139-152
58 Brown, K.D., Bennett, G.N., Lee, F., Schweingruber, M.E. and Yanowsky, C. (1978) J. Mol. Biol. 121,153-177
59 Bennett, G.N. and Yanowsky, C. (1978) J. Mol. Biol. 121,179-192