Structure of open promoter complexes with *Escherichia coli* RNA polymerase as revealed by the DNase I footprinting technique: compilation analysis

O. N. Ozoline* and M. A. Tsyganov[1]

Institute of Cell Biophysics and [1]Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, 142292 Moscow Region, Russian Federation

ABSTRACT

Footprinting data for 33 open promoter complexes with *Escherichia coli* RNA polymerase, as well as 17 ternary complexes with different regulators, have been compiled using a computer program FUTPR. The typical and individual properties of their structural organization are analyzed. Promoters are subgrouped according to the extent of the polymerase contact area. A set of alternative sequence elements that could be responsible for RNA polymerase attachment in different promoter groups is suggested on the basis of their sequence homology near the hyperreactive sites. The model of alternative pathways used for promoter activation is discussed.

INTRODUCTION

DNA-dependent RNA polymerase interacts with promoter DNA by recognizing special signal elements in its structure. Several lines of evidence obtained from individual promoters indicate that canonical hexamers revealed in −35 and −10 regions from the transcription start point are important for recognition and that RNA polymerase makes specific contacts with their bases (1–5). It was suggested that these hexamers are common promoter determinants and that the activity of any promoter should increase as their sequence approaches consensus. However, data from a larger set of promoters display some discrepancies with this hypothesis (4,6–10). Other regions have been proved to be significant for activation (11–20), yet none of them can be found in all promoters. It seems that the absence of any particular promoter-specific element can be compensated for by other structural factors not considered as universal promoter-specific elements (7,9,20–22). The molecular mechanism of polymerase interaction with variable promoter sequences (22) and the architecture of the complexes formed (8,20,21) may therefore be different. Recent advances in footprinting technique make it possible either to generalize or to reject this point of view on the basis of extensive structural data.

DNase I is a traditional and widely used digesting agent. Despite some shortcomings due to sequence-dependent structural

preferences and its relative bulk, this enzyme is an informative tool, as it is sensitive to minor changes in the DNA helix structure. The main goal of this review is to align and compare DNase I footprinting data obtained in different laboratories for RNA polymerase open complexes with various promoters. Since a requirement of positively regulated promoters for activator molecules could be due to some structural factor, footprinting patterns for polymerase contact areas in binary and ternary complexes are analysed comparatively. Some structural properties of the free promoter DNAs are also investigated. The following conclusions are drawn: (i) 19 out of 33 different promoters (closely overlapping, divergent and mutant promoters were not taken into account) have DNase I resistant areas between −28 and −20, suggesting that free promoter DNA often possesses a stiff or narrow minor groove in the spacer region; (ii) the position of the upstream limit of the polymerase footprint varies at least within 2.5 helix turns, suggesting that different types of contacts can be made in this area; (iii) positions −47 ± 2, −38 ± 1 and −26 ± 2 (bottom strand) and −46 ± 2, −33 ± 1 and −24 ± 1 (top strand) are usually either unprotected by polymerase or display increased sensitivity to nuclease, suggesting that despite large variations in the final structure, RNA polymerase at different promoters interacts with the same side of the DNA helix; (iv) the main structural parameters registered by the footprinting technique vary within the same area for RNA polymerase forming binary complexes with promoter and ternary complexes with promoter and regulator, probably indicating that there is no major difference in the molecular mechanism of their complex formation with the enzyme.

Promoters are subgrouped according to the extent of their contact area with polymerase. A number of them exhibiting a high degree of homology with the α-specific rrnBP1 UP element are identified in one of these groups. A set of sequence elements that could be responsible for RNA polymerase attachment in different contact areas is suggested on the basis of sequence homology near the hyperreactive sites. Promoters that are similar in one contact area do not necessarily display similarity in the other. Alternative sequence elements, along with the originally revealed consensus sequences, are probably recognized in different promoters (7,9,18–22). The model of alternative pathways used by RNA polymerase for promoter activation is discussed.

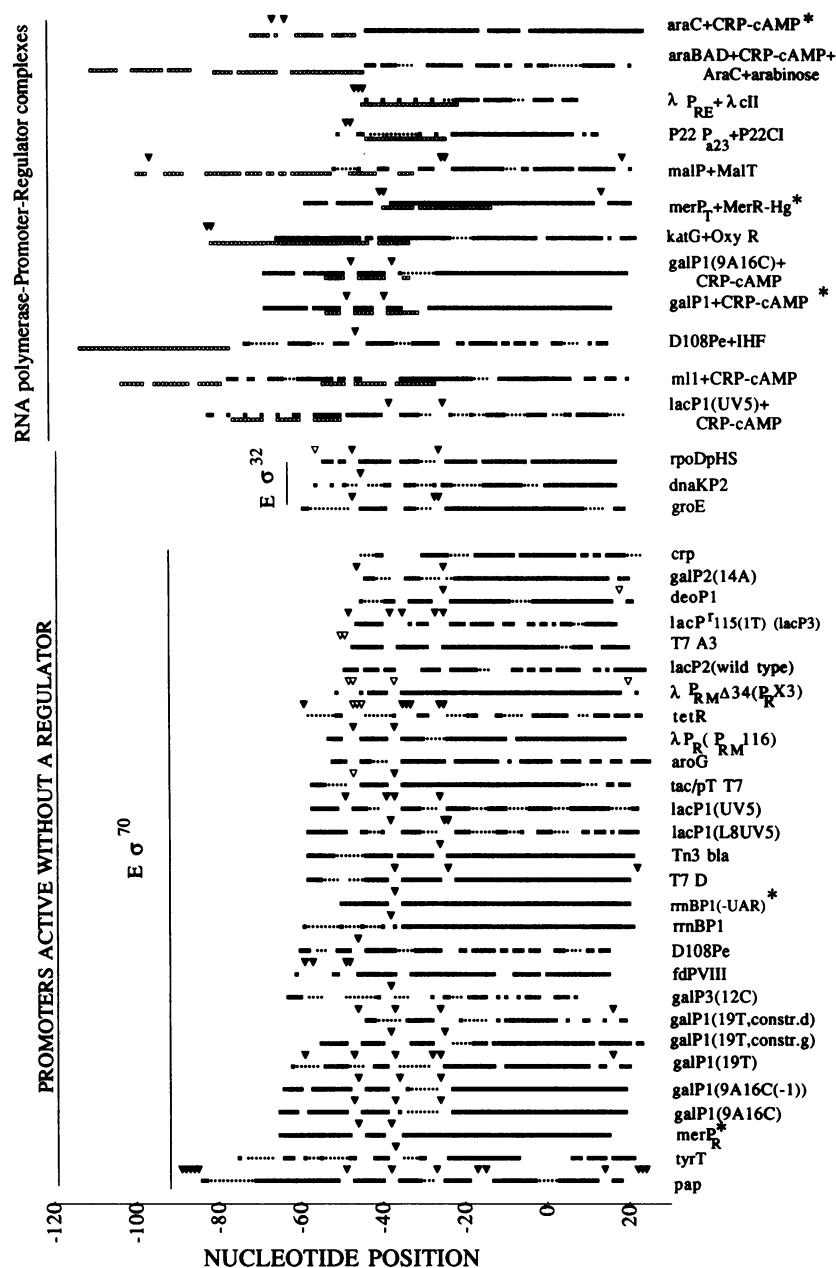* To whom correspondence should be addressed

Figure 1. Schematic presentation of the bottom strand footprinting data. Designations: •, base pairs partially resistant to nuclease without any protein (indicated if at least three consecutive bases are resistant); black line, phosphodiester bonds protected against DNase digestion by RNA polymerase; gaps, unprotected base pairs; dashed line, regions protected by activator molecule; filled triangles, hyperreactive base pairs (∇, base pairs whose hyperreactivity was estimated by FUTPR and is not discussed in the original papers); *, promoter complexes that were not analyzed by FUTPR. References to original papers and discrepancies: *pap* (24); *tyr*T (25); *merP*$_R$ (26); *gal*P1(9A16C) and 9A16C(−1) (27); *gal*P1(19T) (28, agrees with 19); *gal*P1(19T, constructs d and g) (8); *gal*P3 (29); *fd*PVIII (30); D108P$_e$ (31); *rrn*BP1 (+ATP and CTP) (18, agrees with 32); *rrn*BP1(UAR deleted) (+ATP and CTP) (32); T7D (33); Tn3 *bla* (34, agrees with 35), but hyperreactivity at −27 is not registered); *lac*P1(L8UV5) (36), additional hyperreactivity at −48 was registered in (37); *lac*P1(UV5) (38, agrees with 39−41), additional hyperreactivity at +18 registered in (42), hyperreactivity at −49 was not registered in (43); *tac*/pT T7 (44); *aro*G (45); λP$_R$(P$_{RM}$116) (46, agrees with 47 for (P$_R$); λP$_{RM}\Delta34$(P$_R$X3) (46); *tet*R (48); *lac*P2 (36); T7A3 (43); *lac*P^{r}115 (37); *deo*P1 (49); *gal*P2 (28, agrees with 50); *crp* (51); *gro*E, *dna*KP2 and *rpo*DpHS (52); *lac*P1(UV5)+CRP−cAMP (40); *mll*+CRP−cAMP (53); D108P$_e$+IHF (31); *gal*P1+CRP−cAMP (50); *gal*P1(9A16C)+CRP−cAMP (27); *kat*G+OxyR (54), *merP*$_T$+MerR-Hg (26); *mal*P+MalT (55); P22P$_{a23}$+P22 C1 (56); λP$_{RE}$+λcII (57); *ara*BAD+CRP−cAMP+*ara*C+arabinose (CRP−cAMP interacts at −110 + −86, AraC−arabinose interacts at −80 + −44) (58~); *ara*C+CRP-cAMP (58).

COMPILATION PROCEDURE

The compilation includes 30 Eσ^{70} promoter complexes which are active without any regulator. Sixteen of the complexes are at the native promoters, nine with promoters bearing a mutation(s) to activate a wild-type promoter (*lac*L8UV5 and *lac*UV5) or to eliminate the activity of an overlapping promoter [*gal*P1(19T and 9A16C), *gal*P2(14A), *gal*P3(12C), λP$_R$(P$_{RM}$116), λP$_{RM}\Delta34$-(P$_R$X3) and *lac*P^{r}115(1T)], four with mutant promoters, different in overall topology {*gal*P1[9A16C(−1)], *gal*P1(19T, constructs d

and g) and *rrn*BP1(–UAR)} and one with a semisynthetic promoter (*tac*/pT T7). Compilation also includes three Eσ^{32} promoter complexes and 13 (plus three mutant variants) ternary complexes with different positive regulators. Only those promoter complexes which are at least in the transiently open state (confirmed by structural or functional approaches) have been included. If the same promoter was under study in a number of different laboratories all experimental data from those studies were analyzed and the most relevant ones have been presented. Principal discrepancies are indicated in the legends to the figures. Promoter *rrn*BP1, which forms a stable open complex only in the presence of initiating ATP and GTP, is presented as an exception, because of its contacts with the α-subunit of the enzyme in the upstream region. No other initiating polymerase–promoter complexes are included.

To make possible the comparative analysis, published footprinting data were recorded on video camera and analyzed by the computer program FUTPR, written in C. The position and intensity of every electrophoretic band were estimated by the following procedure.

Original autoradiograms were calibrated using a sequencing lane or other size markers. Control and experimental lanes, containing promoter fragments digested by nuclease in the absence and in the presence of polymerase, respectively, were scanned. A quasi-linear approximation for the rate of migration, $V = a - b \times \log L$, where L is the length of DNA fragment used to estimate the position of any band. Coefficients a and b were automatically deduced for subsequent intervals between two marked bands. The accuracy of this procedure was ±1 bp. The program also permits approximation of downstream and upstream from the last marked band, however, in this case accuracy diminishes and reaches 2 for an ~10 bp extrapolation. We avoid the procedure of approximation for more than one helix turn.

Control and experimental densitograms were subtracted (they could also be directly compared) to estimate bases whose reactivity changes upon attachment of the enzyme. Corresponding tabular data obtained for any position and expressed in terms of either 'protection' (solid line), 'no protection' (gap), 'hyperreactivity' (upsidedown triangles) or 'uncertainty' (dots) were transferred to the graphics software. A base pair was considered as protected or hypersensitive if it belongs to the 3′-end of the bottom or to the 5′-end of the top strand of the fragment, the frequency of which appearance correspondingly changes. For example, increased DNase cleavage between base pairs –38 and –37 is considered as hyperreactivity at –37.

Ternary complexes were analyzed in a similar way. In a number of promoters the operator site overlaps the polymerase contact area, thus complicating attribution of protected bases. To standardize the presentation the following strategy was used: (i) the footprinting patterns for a promoter complex formed with the corresponding regulator were estimated; (ii) then the ternary complex with RNA polymerase was analyzed. All changes were considered as induced by the enzyme. If no changes were registered in the overlapping area, but polymerase itself forms contacts with the operator site in the absence of the regulator (*lac*UV5, *gal*P1, *kat*G and λP$_{aQ}$), then common base pairs of the operator were considered as doubly protected. If polymerase alone does not interact with the analyzed promoter, but the footprint of the ternary complex reveals changes both upstream and downstream from the operator site, then the presence of the

polymerase in the overlapping area was disignated by a broken line (P22P$_{a23}$, P22P$_{RE}$, λP$_{RE}$ and *lac*P1).

Troubleshooting

Due to differences in the overall quantity of the DNA loaded onto the gels, control and experimental lanes sometimes differ in intensity. The program permits normalization relative to any unmodified band lying in regions distinct from the contact area. Due to the presence of 3′-hydroxyl groups in the DNase-produced fragments their mobility is slightly greater than for the product of chemical cleavage, used as size markers. FUTPR allows compensation for this difference by a corresponding vertical shift. A more complex situation occurs if the ion electrophoretic front itself displays any curvature. In these cases identical bands in the control and/or experimental lanes have slightly different rates of migration and are also sloped. To overcome this last problem densitograms were taken in the narrow interval from the middle of the lanes.

COMPILATION REMARKS

Initiating this analysis it is important to underline that a slight alteration in the experimental conditions may change observable footprinting patterns. In particular, a decreased enzyme:promoter ratio may lead to the appearance of free promoter DNA, which consequently may show bonds 'available' for cleavage within the contact area. A temperature shift may result in heterogeneity of the complexes analyzed, and so on. Nevertheless, comparative analysis of the data obtained for the same promoter in different laboratories showed us that these variations have usually a quantitative rather than a qualitative character, mostly manifested in the case of cleavable bonds inside the contact area and the precise position of the contact borders. To give an idea of these variations, footprinting patterns obtained by different authors for related promoters [*lac*L8UV5 and *lac*UV5; *gal*P1(19T) and *gal*P1(9A16C)] are included in Figures 1 and 2. One should take into account the possibility of these quantitative variations, but they cannot change the main conclusions made in this compilation. The principal discrepancies usually connected with registration of hyperreactive sites are indicated in the figure legends.

The data were separately summarized for the bottom (sense, Fig. 1) and top (antisense, Fig. 2) strands. Promoter complexes are subdivided into three functional groups and disposed according to the position of the upstream limit of protection. Native and mutant promoters are placed together. Figures 1 and 2 also include seven promoter complexes (marked by asterisks) which are shown to be in strong accord with the schemes presented in the original papers. These complexes were not analyzed by the described procedure, due to absence of the corresponding autoradiograms, their small size or incomplete marking. We, however, decided to include these data because the borders of the protected area, as well as the positions of hyperreactive base pairs, are available for comparison.

Base pairs resistant to DNase I in free promoter DNA

When interacting in the minor groove of the DNA double helix DNase I forms contacts with at least 6 bp along the groove and with the phosphate backbones of both strands (66–68). No contacts exist in the major groove, but on the 3′-side of the
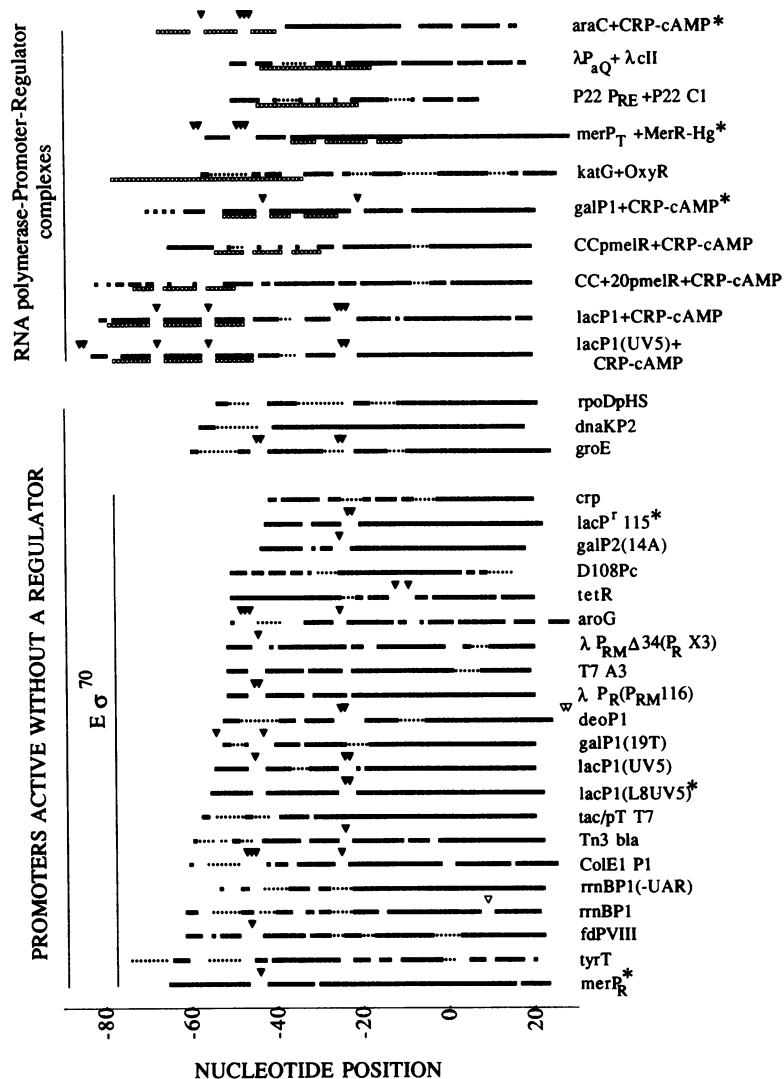
Figure 2. Schematic presentation of the top strand footprinting data. Designations are the same as for Figure 1. References and discrepancies: *rrn*BP1(+ATP and CTP) (32, agrees with 18, but hyperreactivity at +9 is not registered); *rrn*BP1 (–UAR) (32, agrees with 18); *Col*E1 P1 (59; Tn*3 bla* (60, agrees with 35); *lac*P1(L8UV5) (37); *lac*P1(UV5) (43, agrees with 23, 25, 38–41), in (42) additional hyperreactivity is registered at –53; *gal*P1(19T) (28); λP_R(P_RM116) (46); D108P_c (31); *groE* (52, agrees with 61); *lac*P1+CRP–cAMP (62, agrees with 63); CC+pmelR+CRP–cAMP and CC+20pmelR+CRP–cAMP (64); P22 P_RE+P22 C1 (56); λP_aQ+λcII (65); other promoter complexes as in Figure 1.

cleaved bond the protein approaches the DNA backbone again ~5 bp away from the cut site. Access to the major groove will therefore be partially restricted. Several global and local parameters, such as groove width, local bend and DNA stiffness, as well as DNA backbone conformation or the orientation of the scissile phosphodiester bond, may affect the digestion rate (66–69). Decreased activity of DNase at extensive regions could, therefore, be informative about peculiarities of their conformation. Locations of poorly cleaved (in the absence of RNA polymerase) phosphodiester bonds, extending for ≥ 3 bp in different promoters, are presented in Figure 3A. It appears that partially resistant bonds could be found at practically any position throughout the promoter length, although their presence is less apparent immediately upstream from the –10 region. An increase in occurrence, in contrast, is observed upstream from —39 and in the spacer region. A majority of gaps contain A-T runs, which are relatively resistant to nuclease (68).

Enrichment of upstream regions with A-T runs is a widely-discussed property of promoter DNA (7,18). The functional significance of this region is usually explained by its specific functional features and footprinting data are in line with this point of view. Specific base pair composition of the spacer region has been also discussed (8,11,20,21,70). DNase analysis reveals that 19 out of 33 different promoter fragments are relatively resistant to cleavage at –27 and –24 (top strand) and/or–28 and –20 (bottom strand), indicating that free promoter DNA often possesses a stiff or narrow minor groove in the spacer region.

Variation in the upstream limit of the RNA polymerase footprint

A base pair was considered as a border of the contact area if it was the last one that showed a decreased rate of digestion in the
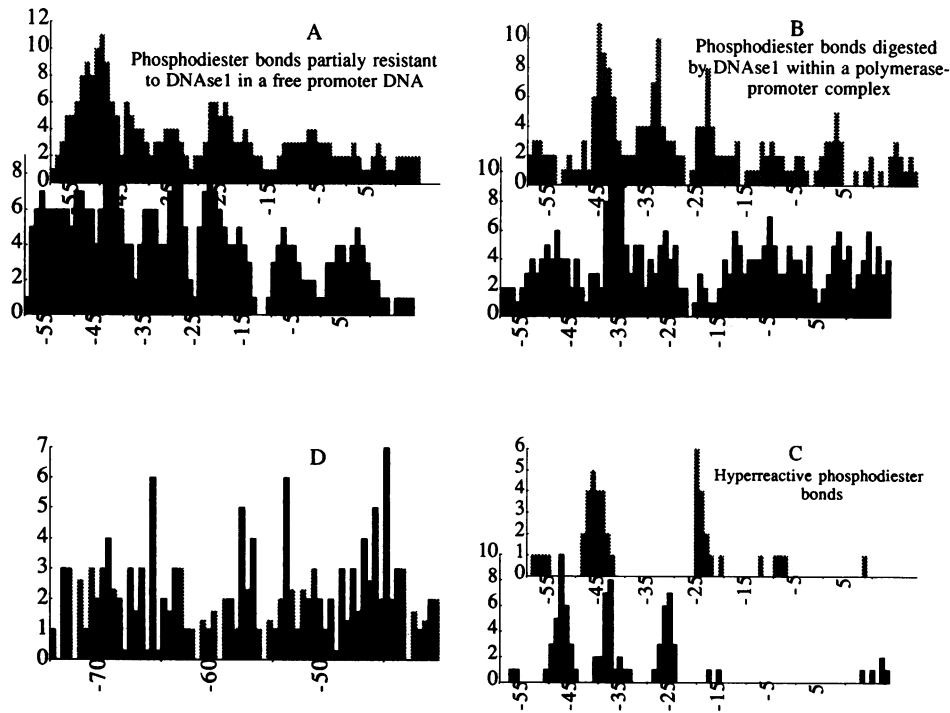
Figure 3. Number of promoters displaying resistance against nuclease digestion for free promoter DNA (A), sensitivity to DNase I within the polymerase contact area (B) and hyperreactivity (C). Footprinting data for the top (half-tone columns) and bottom (black columns) strands are diagrammed. The contribution of mutant promoter complexes if identical to native ones was not taken into account. (D) Distribution of the sequences homologous to the α-specific UP element (black columns) or random sequences (half-tone columns) throughout the length of promoter sequences (from 3). Base pair positions are indicated on the *x*-axes of all plots.

presence of polymerase. It should, however, be taken into account that the borders thus estimated could be inaccurate. In the presence of DNase-resistant sequences at the end of the contact area the limit of the footprint will be underestimated. On the other hand, decreased nuclease digestion at regions near the contact border may be also due to DNA conformational transitions indirectly induced by bound polymerase. In this case the extent of the contact area will be overestimated.

The upstream limit of polymerase contact area varies from −84 (*pap*) to −41 (*crp*), showing no difference for binary and ternary complexes. Visible protection of two strands usually has an offset of 2 bp (the bottom strand shows a more prolonged footprint in the 3′ direction). That is due to the position of the cleaved phosphodiester bond away from the center of the DNase I contact area. Endonuclease requires two unblocked base pairs at the 3′-end and four at the 5′-end from the cut bond (68). In some cases footprinting data reveal equally protected strands (*crp* and *tac*/pT T7) or strands protected with an opposite shift (*rrn*BP1). That could be due to the reasons discussed above. Nevertheless, some promoters (*deo*P1 and T7A3) display an essentially more protected top strand. This phenomena could probably be explained by polymerase interaction with (or a spatial approach to) the top strand sugar–phosphate backbone, thus implying certain peculiarities for the contacts formed at these promoters.

The position of the upstream limit of the polymerase footprint does not correlate with the downstream contact border, which lies near +20 for most of the open enzyme–promoter complexes.

Promoter complexes could probably be subdivided into a number of groups according to the position of their upstream border. The shortest contact region is characteristic for *ara*C,

*ara*BAD, λP$_{RE}$, *crp*, *gal*P2, *lac*P^{r}115 and for the consensus −35 region-containing mutant *gal*P1 (construct d). When aligned according to their upstream limit these promoters display no homologous sequences (Fig. 4). Since in these complexes RNA polymerase reaches only 1–5 bp upstream from the −35 control element, taking into account the mechanism of DNase action it could be suggested that some crucial contacts with the enzyme are formed there. However, no strong homology with the consensus hexamer [except *gal*P1(19T,construct d)] could be observed for this promoter group.

Most promoters have more extended footprints. Ternary complexes with P22P$_{a23}$, *mal*P, λP$_{aQ}$, P22P$_{RE}$, as well as polymerase complexes with D108P$_c$, *deo*P1, T7A3, *lac*P2, λP$_{RM}$, *tet*R, λP$_R$, *aro*G, *rpo*DpHS and truncated *rrn*BP1, could be attributed to the group having the border shifted 9–14(15) bp upstream from the −35 element. In a number of promoters canonical base pairs at −36 and −32 are recognized through the major groove by the σ^{70} helix–turn–helix motif (26,71,72). Since RNA polymerase is supposed to interact in this region with one face of the DNA helix (26,43,48,52,61), additional protection approximately one helix turn upstream implies that polymerase reaches the neighbouring major groove and that the last contacts are probably made with its functional groups. Again, however, none of the alignments permit determination of any obvious consensus sequences for all promoters of this group.

Most other promoter complexes display the limit shifted approximately two helix turns (17–25 bp) from the −35 module. The *tyr*T promoter falls into the third group if the top strand data are taken into account. Its bottom strand is more protected (Fig. 1). Since the cleavage rate of only one bond lying upstream from the nuclease-

```
1.   I GROUP    GCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATaCCCGTTTTTTTGGATGGAG            araBAD+reg
2.               TAGCAAAGTGTGACGCCGTGCAAATAATCAATGTGGACTTTTCTGCCGTGATTATAGACACTTTTGTTACgCGTTTTTGTCATGGCTTTG             araC+reg.
3.               ----------------------CAGCCTCGTCGTGTTTGTTTGCACGAACCATATGTAAGTATTTCCTTaGATAACAAT-----------           λPRE+reg.
4.               AAGGCGACCTGGGTCATGCTGAAGCGAGACACCAGGAGACACAAAGCGAAAGCTATGCTAAAACAGTCAGgATGCTACAGTAATACATTG            crp*
5.               AATTCTTGTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTGTTATACTATGGTTaTTTCATACCATAAGCCTAA            galP2(14A)
6.               CTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGTATTGTGACGGCGATAaCAATTT-----------          lacPr115(1T)
7.               -----------------------N--GAATTCTTGACAGCTGCATGCATCTTTTTTATGGTTATTTCaTACCATAAGCCTAATGGAG            galP1(19T,con.d)

8.   II GROUP   -------------------N--CAAATGTTGCGAGCACTTGCAGTACCTTTGCCTTAGTATTTCCTTCAaGCTTTG--------------        P22 P423+reg.
9.               CCGCAACTCCTCCCTGCCTAATCCCCCGCAGGATGAGGAAGGTCAACATCGAGCCTGGCAAACTAGCGATaACGTTGTGTTGAAAATCTA           malP+reg.
10.              CTGTGCCAATCTTGGCGGTTAGTTCTTTTATTCATATGCATTTCATGTAATCCTTTCAGATTCCCCCGTGaGCTACGCTATAAAGTAGCT           D108Pc*
11.              ---------------TTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCAtTAAAGCTTATCGATGATAA           tetR*
12.              ----------------AATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCtTCCGGCTCGTATGTTGTGT            lacP2
13.              ----------------GCGTTGGCTTTCTTGCGAGTGCTTGTGAGTTCCATATGTGAACATTCCT---------------------          P22 PRE*+reg.
14.              -----------------N--GCGAATGTTGCGAGCACTTGCAGTACCTTTGCCTTAGTATTTCCTTCaaGCT--------------         λPAO*+reg.
15.              ACCGCCAGAGGTAAAATAGCCAACACGCACGGTGTTGATATTTATCCCTTGCGGTGATAGATTTAACGTaTGAGCACAA----------         λPRM Δ34(P,X3)*
16.              AGCGCACCACGGCACATAAGGTGAAACAAAACGGTTGACAACATGAAGTAAACACGGTACGATGTACCACaTGAAACGACAGTGAGTCAC           T7 A3*
17.              GGGTTCTGTCGAGGTAACGCCAGAAACGTTTTATTCGAACATCGATCTCGTCTTGTGTTAGAATTCTAACaTACGGTTGCAACAACGCATC          deoP1*
18.              TCACCGCAAGGGATAAATATTTAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTGATAATGGTTGCaTGTACTAAG----------          λPR(PRM116)
19.              TTCTAGGATTAATCCGTTCATAGTGTAAAACCCCGTTTACACATTCTGACGGAAGATATAGATTGGAAGTaTTGCATTCACTAAGATAAG          aroG
20.              GTAAAACGACGGCCAGTGAATTCCAAATTTCCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCaCTGACACGGAACAACGGCA          rrnBP1(-UAR)*
21.              AGAGCACTATCGTGGTACAAATAATGCTGCCACCCTTGAAAAACTGTCGATGTGGGACGATATAGCAGATaAGAATATTGCTGAGCAAAC         rpoDpHS

22.  III GROUP  ----------CTCCAAATTGTTTTCCATATCGCTTGACTCCGTACATGAGTACGGAAGTAAGGTTACGCTaTCCAATTTCAATTCGAAAG          merPT+reg.
23.              ATCATGATAAATCCTTTCAAATCACTATTGATTTAAAAATGATTCCACACTACATTGCACGATATGCCGTgAGCTACTTTATAGCGTAGC         D108Pe
24.              ATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGaATTGTGAGCGGATAACAAT          lacP1(UV5)
25.              ATGTAAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGaATTGTGAGCGGATAACAAT          lacP1(L8UV5)
26.              CAGTGAGTGAGTCTGCAAAAAAATGAAATTGGGCAGTTGAAACCAGACGTTTCGCCCCTATTACAGACtcaCAACCACATGATGACGGAA          dnaKP2
27.              ----GGATCCACTAATTTATTCTAGATCACACTGCATGCATCTTTTTTATGCTATGGTTATTTCaTACCATAAGCCTAATGGAG            galP1(19T,con.g)
28.              GGTCTGGTCTTTATGTTTAAACTTTAAGATAGGCGTTGACTTGATGGGTCTTTAGGTGTAGGCTTTAGGTgTTGGCTTTAGGATGGACGT          T7 D
29.              TTGTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTTTTATGCTATGGTTATTTCaTACCATAAGCCTAATGGAG         galP1(19T)
30.              GATCGTTTTCACAAAAATGGAGTCCACAGTCTTGACAGGGAAAATGCAGCGGCGTAGCTTTTATGCTGTAGCTTTTATGCTGTATATAA          ColE1 P1*
31.              TTTTGTGCTGATCAGAATTTTTTTTTCTTTTTCCCCCTTGAAAGGGGCGAAGCCATCCCCATTTCTCTGGTCaCCAGCCGGGAAACCACGTA       groE
32.              -------------N------ATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGaTGGGCACCCATCGTTCGTA         tac/pT T7*
33.              ------ACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTgATAAATGCTTCAATAATAT        Tn3 bla*
34.                   TTGTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTGTTCTGCTATAGTTATTTCaTACCATAAGCCTAATGGAG   galP1(9A16C)
35.                   TTGTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTGTTCTGCTATAGTTATTTCaTACCATAAGCCTAATGGAG   galP1(9A16C,-1)
36.              GTTGCGCGGTCAGAAAATTATTTTAAATTTCCTCTTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCaCTGACACGGAACAACGGCA          rrnBP1*
37.              CGACACAATTTATCAGGCGATGATACAAATCTCCGTTGTACTTTGTTTCGCGCTTGGTATAATCGCTGGGgGTCAAAGATGAGTGTTTTA        fdPVIII*
38.              AAGATCTCAACTATCGCATCCGTGGATTAATTCAATTATAACTTCTCTCTAACGCTGTGTATCGTAACGGTAaCACTGTAGAGGGGAG          katG+reg.
39.              TCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTGTTATGCTATGGTTATTTCATACCATAAGCCTaATGGAG------------         galP3(12C)
40.              -----N----------CAGACATGCGCTTGTCCTTTCGAATTGAAATTGGATAGCCTAACCTTACTTCCgTACTCATGTACGGAGTCAA        merPR
41.              ATTACGCAACCAGTTCATTTTTCTCAACGTAACACTTTACAGCGGCGCGTCATTTGATATGATGCGCCCCgCTTCCCCATAAGGGAGCA         tyrT*

42.              GGTACCCGGGGATCAGGTAAATGTGATGTACATCACATGGATCCCCCCTCACTCCTGCCATAATTCTGATa-----------------         CCpmelR*+reg.
43   N-------TAAATGTGATGTACATCACATGGATCCAGATCGATGTTTAAAGATCCCCCCTCACTCCTGCCATAATTCTGATa-----------------      CC+20pmelR*+reg.
44.              TTGTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTGTTATGCTATGGTTATTTCaTACCATAAGCCTAATGGAG         galP1*+reg.
45.              TCCAGATCAGATTCGGGGGATCAGGTAAATGTGATGTACATCACATGGATCCCCCCTCACTCCTGCCATAATTCTGATa-----------------    ML1+reg.
46.              N-AATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGaATTGTGAGCGGATAACAAT  lacP1*+reg.
47. TGCACTGGAAAAATATTTAAACATTTATTCACCTTTTGGCTACTTATTGTTTGAAATCACGGGGCGCACCGTATAATTTGACCgCTTTTTGATGCTTGACTCT     pap
```

Figure 4. Nucleotide sequences of the analyzed promoters subgrouped and aligned according to the position of the upstream limit of the polymerase footprint (indicated by bold italic for either top or bottom strand data). Top or bottom strand data are used to indicate the largest extent of protection. Promoters are marked by asterisks and shifted 2 bases downstream if the top strand data for the border position is presented. Promoters are marked (+reg) if the border position was estimated for ternary complexes. Sequences homologous to the α-specific UP element are double underlined. Sequences homologous to the galP2/P3 upstream module are underlined. Conjectural −35 regions are indicated by bold, +1 positions are marked by lower case letters.

resistant run justifies this prolongation, it probably could be explained by DNA conformational changes indirectly induced by polymerase. Promoter *pap* exhibits an extremely long contact area, comparable with a number of ternary complexes where protein–protein interaction could be responsible for DNA spatial condensation.

The UP element of the promoter *rrn*BP1 (third group) forms contacts with the α-subunit (18), which is required for activation. This UP element interacts with the isolated α-subunit even when transferred into another promoter (19), thus the sequence GAAAATTATTTTAAATTTCCTC contains all signals necessary for recognition. We used another computer program, SSEQ, to search for promoters homologous with the α-specific element. Six promoters of this compilation (*mer*PT, D108Pe, Tn3 *bla*, *gro*E, *pap* and *gal*P1 construct g) have 12–15 coincidences with the *rrn*BP1 UP element near their upstream contact border (Fig. 4). No promoters from the first two groups display so high degree of homology with the analysed sequence.

Approximately 26% of promoter sequences taken from the last compilation (3) show 12–15 matches with *rrn*BP1 UP. Most homologous sequences begin at either −44 ± 1, −53 ± 1, −57 ± 1 (as for *rrn*BP1) or −65 ± 2 (as for D108Pe) (Fig. 3D). Three 22 bp non-promoter sequences with nearly the same A+T content were taken from T7 DNA regions with no promoter-like elements (5'-end at positions 24571, 24800 and 31351) (73) and analysed in the same way as control ones. They show similarity with ~19% of promoters and the distribution of homologous sequences along the promoter length was almost homogenous (Fig. 3D). Therefore, it seems probable that some other promoters can be activated through α-specific interaction with a *rrn*BP1-like UP element, however, other promoters which have prolonged contact with the enzyme have an essentially different sequence in this region.

The *gal* control region contains three overlapping promoters. Without CRP–cAMP, RNA polymerase preferentially interacts with *gal*P2, in which the transcription start point is shifted 5 bp upstream of the *gal*P1 start site. Substitution of C for −12T inactivates both

P1 and P2 and transcription initiates from *gal*P3, which is 14 bp downstream of the *gal*P1 start point. To our surprise, the footprinting analysis shows that, although essentially different in overall topology, the P2 and P3 promoter complexes have the same upstream limit of protection (Fig. 4). The sequence in this region, ATTCCATGTCACACTTTTCGCA (partly overlapping the CRP binding site) was analysed in the same way as the α-specific element. It appears that homologous sequences could be found in *ara*BAD, *ara*C and *crp* from the first promoter group and in *tyr*T, *Col*E1P1 and *lac*L8UV5 from the third promoter group (Fig. 4). Most of the promoters (except *gal*P1 variants and *tyr*T) contain these sequences near their upstream limit of the polymerase footprint. Approximately 18% of promoters from the Lisser and Margalit compilation have 12–15 matches with the *gal*P2/P3 element. They preferentially begin at positions –41 ± 1 (as in *crp* and *gal*P2), –45 ± 1(as in *ara*C and *ara*BAD) and –49 (as in *tyr*T). Control sequences (5'-end at positions 30838, 31013 and 31339 of T7 DNA) have 12–14 matches with ~12% of promoters. Perhaps the *gal*P2/P3 UP sequences also contain some elements readily recognized by polymerase. Since both these promoters are weak, this interaction could be unfavourable and regulatory molecules could be needed for activation. No promoter of the second group showed homology with either of the UP elements analysed.

It could be concluded that promoters subdivided according to their upstream contact border share no common sequence motifs in any of individual groups. Any classification could only be made by taking into account additional data. Different elements are probably recognized by RNA polymerase in upstream regions and variable types of contacts are formed.

Phosphodiester bonds accessible to DNase I digestion within the polymerase–promoter contact area and positions displaying hyperreactivity

Some phosphodiester bonds within the contact area with polymerase remain exposed to nuclease digestion. Promoters display the highest degree of variability in the position and reactivity of these scissile bonds (Figs 1, 2 and 3B). This variability may, at least partially, be due to differences in the relative contrast of the original autoradiograms and the presence of free DNA in the experimental samples. However, none of these reasons can explain the non-random location of the scissile bonds, which preferentially remain at positions –47 ± 2, –37 ± 2, –26 (and –25) ± 1 for the bottom and –45 ± 2, –33, –34 and –25 ± 1 for the top strand respectively. Their approximately one helix turn phased disposition probably indicates that in different promoters RNA polymerase generally interacts with the same face of the DNA double helix. However, the total absence of scissile bonds is observed at only a few positions. In the region from —20 to +20 unprotected base pairs display almost no preference in location and the bottom strand as a rule possesses more scissile bonds.

The most documented characteristic of enzyme–promoter complexes is the presence and location of sites which are more sensitive to nuclease upon polymerase attachment. Only one enzyme–promoter complex of this compilation (*crp*) displays no hypersensitivity for both strands. In some cases increased sensitivity to nuclease is registered outside the contact area and could be explained by cooperative changes induced by enzyme in the neighbouring DNA conformation [*gal*P2, *lac*P^{r}115, T7A3, *tet*R, T7D, *gal*P1(19T) and *pap*]. The majority of hyperreactive

sites appear within the contact region (Figs 1, 2 and 3C). Fourteen out of 39 enzyme–promoter complexes (mutants and the same promoters forming ternary complexes are not taken into account) have at least one hypersensitive site on the bottom strand just upstream from the –35 module, –38 and –37 being preferentially cleaved. Taking into account the offset of DNase cleavage for the two strands, hyperreactive sites on the top strand would be expected near position –35. However, the top strand never shows hyperreactivity in this region. Only the *cat*P ternary complex with polymerase and CRP–cAMP could be an exception to this rule, since increased sensitivity to nuclease is registered at –36; unfortunately, it is not clear which strand autoradiogram is presented in the paper (74). Since the top strand of many promoters possesses scissile (not hyperreacvtive) bonds at –33 and –34 (Fig. 3B), it could be suggested that hyperreactivity in the region just upstream of the –35 control element is induced by tight polymerase contacts within the major groove restricting DNase access in the 5' direction.

Twenty three and 14 promoter complexes have hypersensitivity in the narrow areas near positions –47 and –26 (–25) (–46 and –25 for the top strand) respectively. In these regions hypersensitivity appears with nearly equal probability for the top and bottom strands. No differences in the position of hyperreactive bonds or probability of their appearance can be registered for binary and ternary complexes. Enzyme–promoter complexes do not usually have hyperreactive bonds in the region from —20 to +10. This is in line with the model suggesting that in this region RNA polymerase covers a greater surface of the helix (26,46). Again, this is ruled out by the promoters *pap* and *tet*R (Figs 1 and 2).

At least eight promoter complexes are hyperreactive downstream of the transcription start point. The locations of these reactive bonds are not phased with the helical turns.

Promoter subgrouping according to the sequence motifs surrounding the hyperreactive sites

Because of DNase I binding the minor groove of the DNA undergoes local conformational transitions. It opens by ~3 Å and the duplex bends towards the major groove by ~20° (66–68). DNase by itself remains nearly unchanged, providing a rigid framework to which the DNA substrate has to adapt (68). Any DNA segments stabilized by RNA polymerase in a conformation favourable for cleavage would exhibit an enhanced sensitivity to nuclease. Thus the appearance of hyperreactive sites in the promoter DNA indicates that some tight contacts locally modifying minor groove structure are made by the polymerase. A comparison of the sequences in the region adjacent to the reactive bond may provide information on the mechanism of RNA polymerase–promoter interaction.

Promoter segments surrounding hyperreactive sites were consecutively aligned. It appears that some sequence motifs could be found within one helix turn region from the cut site (Fig. 5). Promoters exhibiting similarity upstream of the hyperreactive bond are not necessary similar in the downstream area, suggesting that sequence-specific contacts with polymerase, if any, can be made independently. Promoters displaying no enhanced sensitivity to DNase in any of the analyzed areas usually show little or no homology with characteristic sequences. Some sequence motifs revealed for upstream regions are parts of the *rrn*BP1 and *gal*P2/P3 UP elements. The main sequence motifs in the –25 and–37 contact area resemble TTGACA or its elements. The lowest degree of

UPSTREAM FROM ~-47		-47+ -37 CONTACT AREA		-37+ -25 CONTACT AREA		DOWNSTREAM FROM ~-26	
---ATTC---YW		----WWTTCC--		--TT-ACA----		YKZRTCTTK-KK	
TTATTCACCTTT	47	31 TTCTTTTTCCCC	31	32 CTGTTGACAATTAAT	32	29 CGCATCTTTTTT	
ACGATTCCACTAA	5	5 AATTTATTCCAT	5	30 TCTTGACAGGGAAAAT	30	34 CGCATCTTTGTTC	
----TTCATACAC	11	39 ACACTTTTCGCA	39	7 TCTTGACAGCTG	7	5 TTCGCATCTTTGTTA	
CCGTTCATAGTG	19	36 TTTAAATTTCCT	36	41 ACTTTACAGCGGC	41	17 TCTCGTCTTGTGT	
		6 ACACTTTATGCTTC	6	24 GGCTTTACACTTTA	24	28 ATGGGTCTTTAGG	
--AAAT-----WY				25 GCTTTACACTTTATG	25		
AAAAATGGAAGTCC	30	-CACC-YR---		17 TTCGAACATCGAT	17	11 CAATTTAACTGTGATAAA	
CCAAATTGTTTTC	22	18 ACACCGTGCGT	18	19 CGTTTACACATTC	19	6 TATGTTGTGTGGTAT	
TAAAATAGCCAAC	15	24 AGGCACCCCAGGC	24	5 ATGTCACACTTT	5	24 ATGCTTCCGGCTCGT	
ATAAATATTTAAC	18	25 AGGCACCCCAGGG	25	36 TCTTGTCAGGCC	36	25 TGCTTCCGGCTCGT	
CAGAATTTTTTTTC	31	30 TCCACAGTCTTG	30			27 ATCTTTTTTATGCT	
CTTTCAAATCACTATT	23	41 CTCAACGTAACAC	41	--YWZACT------			
CAAAAAAATGA	26	29 TATTCCATGTC	29,34	18 TGTTGACTATTTTA	18	47 CACGGGGGCGCAC	
GGTACAAATAA	21			28 CGTTGACTTGATGG	28	19 CTGACGGAAGATAT	
		-----GC----Y--		22 CGCTTGACTCCGTA	22	30 TGCAGCGGCGTAGC	
---CAC--WW--		11 CACGGTGCCTGACTG	11	29 CACACTTTTCGC	29	9 ACATCGAGCCTGG	
GCTCACTCATTA	24	21 ATGCTGCCACCCTT	21	34 CACACTTTTCGC	34	7 GCATGCATCTTT	
TTCCACTAATTT	29,34	47 TTGGCTACTTAT	47	27 TCACACTGCATGCA	27	31 GGCGAAGCCATCCC	
CGGCACATAAGGT	16	15 AACACGCACGGT	15			21 TGTCGATGTGGG	
ATTCACCTTTTG	47			-YY-TTGAA---		33 GTATCCGCTCAT	
		--ATTGW--W		40 CCTTTCGAATTGA	40		
ATCAGGCGATGAT	37	23 CTATTGATTTA	23	31 CCCCTTGAAGGGG	31	No hyperreactivity in	
GCCGTGCAAATAA	2	26 AAATTGGGCAG	26	21 ACCCTTGAAAAACT	21	-26 (both strands)	
CCCCAGGCTTTAC	6	32 CTGAAATGAGCT	32	47 TTGTTTGAAATC	47		
---CAGACATG	40	2 ATAATCAATGTGGA	2	33 TACATTCAAATATG	33	4 AGCGAAAGCTAT	
		40 GCGCTTGTCCT	40	39 ATCTTTGTTATGC	39	15 ATCCCTTGCGGT	
No hyperreactivity in		27 TTTATTCTAGAT	27	15 TGTTGATATTTA	15	16 GAAGTAAACACG	
-47 (both strands)						18 TTACCTCTGGCG	
		---ACAR-----		6 CTTCCGGCTCGTAT	6	32 TAATCATCGGCT	
TCATGCTGAAGC	4	30 TCCACAGTCTTG	30	9 GAGGAAGGTCAAC	9	36 CGGAATAACTCC	
GTAACGCCAGAA	17	37 GATACAAATCTC	37	11 CTGCGTTAGCAATTT	11	37 GTTTCGCGCTTG	
TTTGTTTATTTT	33	16 GTGAAACAAAACG	16			40 AAATTGGATAGC	
AAAATTATTTA	36	19 GTGTAAAACCCC	19	No Hyperreactivity in		41 GCGCGTCATTTG	
TTCATTTTTCTC	41	27 TTTATTCTAGAT	27	-37 and -26			
		28 TTTAAGATAGGC	28				
		22 TTCCATATCGCT	22	4 AGGAGACACAAAG	4		
				16 GTTGACAACATGA	16		
		No hyperreactivity in -47		37 CGTTGTACTTTGT	37		
		and -37 (both strands					
		4 AGCGAGACACCA	4				
		17 GAAACGTTTTAT	17				
		33 TTTCTAAATACA	33				

Figure 5. Parts of the promoter sequences (numbered according to Fig. 4) neighboring hyperreactive base pairs (underlined or double underlined for top or bottom strand data respectively). Promoters are subgrouped according to common sequence motifs in any of the probable contact areas. Bold indicates revealed sequence motifs in the promoters with no hyperreactivity on both strands of the analysed region. W means A or T; Z means G or C; R means A or G; Y means T or C; K means G or T.

interrelation is observed downstream of the –25 hyperreactive area. Reactivity in this region is usually explained by a stressed conformation of DNA induced by polymerase contacts with the –35 and –10 sequences and is considered a characteristic of long spacer promoters. However, many promoters that have optimal (*Col*E1 P1, T7D and Tn*3 bla*) or 16 bp (*gal*P2, *pap* and *lac*Pr115) spacers are hyperreactive near –25. Additional sequence-specific contacts made by the enzyme in the spacer region could be a reason for a DNA stressed conformation, thus explaining the altered reactivity to DNase and suggesting that these structural elements, as well as some others revealed on the basis of the compilation analysis, could serve as ununiversal targets for interaction with RNA polymerase.

CONCLUSION

The global structural characteristics observed from the compilation analysis of footprinting data (the extent of protected area and the presence and location of scissile and hypersensitive sites) vary essentially in the same area for formation of binary and ternary complexes of RNA polymerase with promoters. No classification based on the structural data allows determination of common sequence homologies for any of the individual groups. Alternative sequence elements, along with the originally revealed consensus sequences, can probably be recognized in different promoters (7,9,18–22). This is in line with the model of alternative pathways proposed for promoter activation (22). The model suggests that at every step of complex formation the promoter recognition centre of RNA polymerase can identify a set of promoter-specific elements that can be basically different in nature and position. In a given promoter, polymerase preferentially utilizes only one of them. Specific conformational transitions accompany the interaction and create prerequisites for the next stage, at which some alternative possibilities for the enzyme to form further contacts also exist. Alterations in the recognition centre of the enzyme or in the structure of the promoter by regulatory molecules or mutations may switch the process of complex formation from one path to another. For example, a 1 bp mutation (19T) in the spacer region of the overlapping P1 and P2 promoters from the *gal* control region switches the process of complex formation from the *gal*P2 to the *gal*P1 pathway, with essential differences in the final structure formed from practically the same base pair sequence. Further

analysis revealed differences in the organization of open complexes at *gal*P1(19T) according to whether or not a 'consensus-like' –35 sequence was present (Fig. 1) (8). The model suggests that there may be several ways for even one promoter to be activated. In particular, two forms of open complexes differing in electrophoretic mobility and the ability to switch from the cyclic abortive to the productive reaction have been proved for *lac*UV5 (38) and observed for T7D (22). An advantage of the model is its ability to explain how a large variety of different promoters can be activated using a limited set of structural elements of a single promoter recognition centre. Moreover, the process of complex formation may be described in the same terms for both constitutive and regulated promoters.

The model of alternative pathways implies the necessity of finding non-universal promoter-specific elements involved in activation of different types of promoters. An attempt to identify some of them on the basis of structural data is made in this paper. Biochemical and compilation analysis of the data obtained for subsequent stages of complex formation at different promoters could be a useful tool for further progress.

ACKNOWLEDGEMENTS

REFERENCES

1 Hawley,D.K. and McClure,W.R. (1983) *Nucleic Acids Res.*, **11**, 2237–2255.
2 Harley,B. and Reynolds,R.P. (1987) *Nucleic Acids Res.*, **15**, 2343–2361.
3 Lisser,S. and Margalit,H (1993) *Nucleic Acids Res.*, **21**, 1507–1516.
4 Kobayashi,M., Nagata,K. and Ishihama,A (1990) *Nucleic Acids Res.*, **18**, 7367–7372.
5 Szoke,P.A., Allen,T.L. and Haseth,P.L. (1987) *Biochemistry*, **26**, 6188–6194.
6 Horwitz,A.H., Morandi,C. and Wilcox,G. (1980) *J. Bacteriol.*, **142**, 659–667.
7 Deuschle,U., Kammerer,W., Gentz,G. and Bujard,H. (1986) *EMBO J.*, **5**, 2987–2994.
8 Chan,B., Spassky,A. and Busby,S. (1990) *Biochem. J.*, **270**, 141–148.
9 Keilty,S. and Rosenberg,M. (1987) *J. Biol. Chem.*, **262**, 6389–6395.
10 Brunner,M. and Bujard,H. (1987) *EMBO J.*, **6**, 3139–3145.
11 Travers A.A. (1987) *CRC Crit. Rev. Biochem.*, **22**, 181–219.
12 Bossi,L. and Smith,D.M. (1984) *Cell*, **39**, 643–652.
13 McAllister,C.F. and Achberger,E.C. (1988) *J. Biol. Chem.*, **263**, 11743–11749.
14 Bertrand-Burgraff,E., Dunand,J., Fuchs,R.P.P. and Lefevre,J.F. (1990) *EMBO J.*, **9**, 2265–2271.
15 Gourse,R.L., deBoer,H.A. and Nomura,M. (1986) *Cell*, **44**, 197–205.
16 O'Neill,M.C. (1989) *J. Biol. Chem.*, **264**, 5522–5530.
17 OÖNeill,M.C., Chiafari,F. (1989) *J. Biol. Chem.*, **264**, 5531–5534.
18 Ross,W., Gosink,K.K., Salomon,J., Igarashi,K., Zou,C., Ishihama,A., Severinov,K. and Gourse,R.L. (1993) *Science*, **262**, 1407–1413.
19 Attey,A., Belyaeva,T., Savery,N., Hoggett,J., Fujita,N., Ishihama,A. and Busby,S. (1994) *Nucleic Acids Res.*, **22**, 4375–4380.
20 Belyaeva,T., Griffiths,L., Minchin,S., Cole,J. (1993) *Biochem. J.*, **296**, 851–857
21 Chan,B. and Busby,S. (1989) *Gene*, **84**, 227–236.
22 Ozoline,O.N., Uteshev,T.A., Masulis,I.S. and Kamzolova S.G. (1993) *Biochim. Biophys. Acta*, **1172**, 251–261.
23 Gourse,R.L. (1988) *Nucleic Acids Res.*, **16**, 9789–9809.
24 Kanazawa,H., Mabuchi,K. and Futai,M. (1982) *Biochem. Biophys. Res. Commun.*, **107**, 568–576.
25 Travers,A.A., Lamond,A.I., Mace,H.A.F. and Berman,M.L. (1983) *Cell*, **35**, 265–273.
26 O'Halloran,T.V., Frantz,B., Shin,M.K., Ralston,D.M. and Jeffrey,J.G. (1989) *Cell*, **56**, 119–129.
27 Lavigne,M, Kolb,A. and Buc,H. (1992) *Biochemistry*, **31**, 9647–9656.
28 Busby,S., Spassky,A. and Chan,B. *Gene*, **53**, 145–152.
29 Ponnambalam,S., Spassky,A. and Busby,S. (1987) *FEBS Lett.*, **219**, 189–196.
30 Hofer,B., Muller,D. and Koster,H. (1985) *Nucleic Acids Res.*, **16**, 5995–6013.
31 Kukolj,G. and Dubow,M.S. (1992) *J. Biol. Chem.*, **267**, 17827–17835.
32 Newlands,J.T., Ross,W., Gosink,K.K. and Gourse,R.L. (1991) *J. Mol. Biol.*, **220**, 569–583.
33 Ozoline,O.N., Masulis,I.S., Chasov,V.V., Demina,N.N. and Kamzolova,S.G. (1995) *Russian Chem. Bull.*, **44**, 1321–1326 (in English).
34 Duval-Valentin,G. and Reiss,C. (1990) *Mol. Microbiol.*, **4**, 1465–1475.
35 Russel,D.R. and Bennett,G.N. (1981) *Nucleic Acids Res.*, **9**, 2517–2533.
36 Amouyal,M. and Buc,H. (1987) *J. Mol. Biol.*, **195**, 795–808.
37 Peterson,M.L. and Reznikoff,W.S. (1985) *J. Mol. Biol.*, **185**, 525–533.
38 Straney,D.C. and Crothers,D.M. (1985) *Cell*, **43**, 449–459.
39 Schmitz,A. and Galas,D.J. (1979) *Nucleic Acids Res.*, **6**, 111–137.
40 Kolb,A., Igarashi,K., Ishihama,A., Lavigne,M., Buckle,M. and Buc,H. (1993) *Nucleic Acids Res.*, **21**, 319–326.
41 Straney,D.C. and Crothers,D.M. (1987) *J. Mol. Biol.*, **193**, 267–278.
42 Spassky,A., Kirkegaard,K. and Buc,H. (1985) *Biochemistry*, **24**, 2723–2731.
43 Kovacic, R.T. (1987) *J. Biol. Chem.*, **262**, 13654–13661.
44 Krummel,B. and Chamberlin,M.J. (1989) *Biochemistry*, **28**, 7829–7842.
45 Baseggio,N., Davies,W.D. and Davidson,B.E. (1990) *J. Bacteriol.*, **172**, 2547–2557.
46 Woody,S.T., Fong,R.S.-C. and Gussin,G.N. (1993) *J. Mol. Biol.*, **229**, 37–51.
47 Hersberger,P.A., de Haseth,P.L. (1991) *J. Mol. Biol.*, **222**, 479–494.
48 Duval-Valentin,G. and Ehrlich,R. (1986) *Nucleic Acids Res.*, **14**, 1967–1983.
49 Jeppesen,C., Buchardt,O., Henriksen,U. and Nielsen,P.E. (1988) *Nucleic Acids Res.*, **16**, 5755–5770.
50 Taniguchi,T. and de Crombrugghe,B. (1983) *Nucleic Acids Res.*, **11**, 5165–5180.
51 Aiba,H. (1983) *Cell.*, **32**, 141–149.
52 Mecsas,J., Cowing,D.W. and Gross,C.A. (1991) *J. Mol. Biol.*, **220**, 585–597.
53 Busby,S., West,D., Lawes,M., Webster,C., Ishihama,A. and Kolb,A. (1994) *J. Mol. Biol.*, **241**, 341–352.
54 Tao,K., Fujita,N. and Ishihama,A. (1993) *Mol. Microbiol.*, **7**, 859–864.
55 Danot,O. and Raibaud,O. (1994) *Mol. Microbiol.*, **14**, 335–346.
56 Ho,Y.S., Pfarr,D., Strickler,J. and Rosenberg,M. (1992) *J. Biol. Chem.*, **267**, 14388–14397.
57 Ho,Y.-S., Wulff,D.L. and Rosenberg,M. (1983) *Nature*, **304**, 703–708.
58 Lee,N.L., Gielow,W.O. and Wallace,R.G. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 752–756.
59 Chan,P.T. and Lebowitz,J. (1983) *Nucleic Acids Res.*, **11**, 1099–1116.
60 Wishart,W.L., Machida,C., Ohtsubo,H. and Ohtsubo,E. (1983) *Gene*, **24**, 99–113.
61 Cowing,D.W., Mecsas,J., Record,M.T.,Jr and Gross,C. (1989) *J. Mol. Biol.*, **210**, 521–530.
62 Ren,Y.L., Garges,S., Adhya,S. and Krakow,J.S. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 4138–4142.
63 Straney,D.C., Straney,S.B. and Crothers,D.M. (1989) *J. Mol. Biol.*, **206**, 41–57.
64 Bell,A., Gaston,K., Williams,R., Chapman,K., Kolb,A., Buc,H., Minchin,S., Williams,J. and Busby,S. (1990) *Nucleic Acids Res.*, **18**, 7243–7250.
65 Ho,Y.S. and Rosenberg,M. (1985) *J. Biol. Chem.*, **260**, 11838–11844.
66 Suck,D., Lahm,A. and Oefner,C. (1988) *Nature*, **332**, 464–468.
67 Suck,D. and Oefner,C. (1986) *Nature*, **321**, 620–625.
68 Lahm,A. and Suck,D. (1991) *J. Mol. Biol.*, **222**, 645–667.
69 Drew,H.R. and Travers,A.A. (1984) *Cell*, **37**, 491–502.
70 Beutel,B.A. and Record,M.T.,Jr (1990) *Nucleic Acids Res.*, **18**, 3597–3603.
71 Gardella,T., Moyule,H. and Susskind,M. (1989) *J. Mol. Biol.*, **206**, 579–590.
72 Siegele,D.A., Hu,J.C., Walter,W. and Gross,C.A. (1989) *J. Mol. Biol.*, **206**, 591–604.
73 Melancon,P., Burgess,R.R. and Record,M.T. (1982) *Biochemistry*, **21**, 4318–4331.
74 Le Grice,S.F.J. and Matzura,H. (1981) *J. Mol. Biol.*, **150**, 185–196.