Promoter Recognition by *Escherichia coli* RNA Polymerase: Effects of Base Substitutions in the -10 and -35 Regions†

Patricia A. Szoke, Todd L. Allen, and Pieter L. deHaseth*

Department of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106

Received February 27, 1987; Revised Manuscript Received May 15, 1987

ABSTRACT: We have constructed the P_{RM} promoter of phage λ and eight variants, which represent intermediates in the conversion of this promoter to one that has complete homology to the consensus sequences in the -10 and -35 regions. The in vivo activity of these promoters was determined from the β-galactosidase or galactokinase activities in cells harboring plasmids, in which the cloned promoters were driving the expression of these genes. Additionally, the kinetics of the interaction of *Escherichia coli* RNA polymerase with the same series of promoters was measured as a function of RNA polymerase concentration. This allowed the overall rate of functional or open complex formation to be dissected into the equilibrium constant for binding of the polymerase to form a closed promoter complex and the rate of subsequent isomerization to yield the open complex. The following conclusions can be drawn from the data presented: (1) The consensus sequence is optimal for promoter function both in vivo and in vitro. (2) Alterations of the -10 and -35 regions have similar effects on the kinetics of RNA polymerase binding in vitro; with one exception, the same holds for promoter activity in vivo. (3) The in vitro rate of RNA polymerase binding to a promoter is solely determined by the number of positions at which its -10 and -35 regions match the consensus promoter sequence. The functional importance of a match does not appear to be determined by the sequence conservation at the particular position. (4) The extent to which a particular base change affects the kinetic parameters depends on the sequence of the promoter into which it is introduced.

A vast body of data [reviewed in von Hippel et al. (1984) and McClure (1985)] has implicated two upstream regions as being crucial to the function of *Escherichia coli* promoters. These are referred to as the -10 and -35 regions to indicate their approximate distance from the start site of transcription. From the comparison of a large number of promoters (Hawley & McClure, 1983) the consensus sequences TATAAT and TTGACA have emerged for these regions, respectively. No naturally occurring *E. coli* promoter has yet been found that has consensus sequences in both regions. Mutations that increase the homology of a promoter to the above sequences also enhance promoter activity, in vitro as well as in vivo (Youderian et al., 1982; von Hippel et al., 1984;, McClure et al., 1985). Promoter constructs with consensus -35 and -10 regions are clearly better in vitro than otherwise similar constructs with sequence differences in these regions (Mulligan et al., 1984; Brosius et al., 1984), but there is presently no agreement whether this holds in vivo (Inouye & Inouye, 1985; Mandecki et al., 1985).

The sequence of events taking place when RNA polymerase interacts with a promoter can be broadly divided into promoter binding of the enzyme to form an inactive or "closed complex", followed by an isomerization to yield the active or "open" complex (Chamberlin, 1974; McClure, 1980). The overall kinetics of open complex formation can be dissected into the contributions of each of these processes by analyzing data obtained at several RNA polymerase concentrations (McClure, 1980). It has been found that mutations at either the -10 or the -35 regions can affect both the binding and the isomerization steps (Hawley & McClure, 1982), but no systematic study has been performed on the relative contribution of mutations in the two regions to each of the kinetic steps.

The relative in vivo activities of promoters is apparently determined by their relative rates of open complex formation (Maquat & Reznikof, 1978; Stefano et al., 1980). While it seems reasonable that the relevant kinetic parameter would be the second order rate constant for the overall process of open complex formation, no experimental studies have been undertaken that directly bear on this question. In this paper we attempt to address the above problems through a systematic study of the effects of base pair changes in the -10 and -35 regions on promoter function in vivo as well as in vitro.

Materials and Methods

Chemicals and Enzymes. ATP, UTP and UpA were from Sigma. 1-Naphthylamine-5-sulfonic acid (AmNS) was a gift from Dr. Yarbrough and also purchased from Fluka. 1-Ethyl-3-[(dimethylamino)propyl]carbodiimide (EDC) was from Pierce. Chemicals for DNA synthesis were from Applied Biosystems and American BioNetics. Restriction enzymes, the Klenow fragment of DNA polymerase I, and DNA ligase were from New England Biolabs and Boehringer. *E. coli* RNA polymerase was purified and characterized as previously described; our preparation was 40% active (Auble et al., 1986; Bruzik et al., 1987).

DNA. Promoter variants were assembled from chemically synthesized DNA fragments as previously described (Auble et al., 1986). All variants were initially cloned into the vector pRZ5202 (Munson, 1983; Auble et al., 1986). Sequences were verified by dideoxy DNA sequencing as described (Auble et al., 1986). Large-scale isolation of plasmids with the cloned promoters was by standard methods (Maniatis et al., 1982), except that the amplification step was omitted. This modification led to improved yields of the plasmids containing the stronger promoters. All promoters were subsequently sub-

† This work was supported by National Institutes of Health Grant GM31808 and a grant from the American Cancer Society, Cuyahoga County Division.

* Address correspondence to this author.

0006-2960/87/0426-6188$01.50/0

cloned into the *Eco*RI and *Bam*HI sites of the vector pTA03, a promoter probe vector with galactokinase (gal K) as the reporter gene. pTA03 was obtained from vector pKB2000, a derivative of pK01 (McKenney et al., 1982), which contains a region of bacterial DNA upstream of the galactokinase gene. An *Eco*RI site flanking the bacterial DNA was removed by filling in the overhanging ends of the *Eco*RI-cut DNA with the Klenow fragment of DNA polymerase I, followed by religation of the vector. A polylinker containing *Eco*RI, *Pst*I, and *Bam*HI sites was introduced into the *Hin*dIII site immediately upstream of the gal K gene to generate the pTA03 vector. This vector is more useful than pRZ5202 for the assay of strong promoters, since its copy number is less sensitive to the strength of the cloned promoters (results not shown; see below).

Abortive Initiation Assay. Detection of open complex formation was by the abortive initiation assay (McClure et al., 1980) as modified by Bertrand-Burggraf and Daune (1984). The promoter- and polymerase-dependent synthesis of UpApU from UpA and AmNS-UTP [synthesized as described by Yarbrough et al. (1979)] was followed by monitoring the concomitant increase in AmNS fluorescence (excitation at 350 nm, emission at 460 nm). The solutions were contained in a thermostated cuvette, and the measurements were carried out on an Aminco Bowman spectrofluorometer with internal correction for fluctuations in the intensity of the exciting light source. The reactions were carried out in 100 μL (final volume) containing 0.03 M tris(hydroxymethyl)-aminomethane hydrochloride (Tris-HCl), pH 8.0, 0.1 M KCl, 3 mM $MgCl_2$, 0.01 M dithiothreitol (DTT), 0.1 mM ethylenediaminetetraacetic acid (EDTA), 100 μg/mL bovine serum albumin (BSA), 0.5 mM UpA, 100 μM AmNS-UTP, 23 μL of RNA polymerase in storage buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA, 0.1 mM DTT, 0.1 M NaCl, 50% glycerol) to give the desired final concentration of RNA polymerase, and 2.0 nM of PstI-linearized pRZ5202 DNA harboring the desired cloned promoter. The reaction was initiated by addition of the DNA to the mixture containing all other components. Following rapid mixing by allowing five bubbles to rise through the solution, the fluorescence of the solution was read at regular time intervals to obtain a lag curve (McClure, 1980). Control experiments had indicated that most, but not all, of the ensuing fluorescence signal was dependent on the presence of the cloned promoters. The additional signal was defined by using the pRZ5202 vector at each RNA polymerase concentration and subtracted from the experimental values. From these corrected values the lag times τ were obtained by computer fitting (McClure, 1980; Auble et al., 1986). The useful range of this technique as used by us is delimited on the one hand by the rate of mixing (estimated to be 5 s), on the other hand by the stability of the fluorescent signal. Incubations of more than 1 h were avoided.

In Vivo Assays. The assays for β-galactosidase activity were carried out at 28 °C on lysed cells containing promoters cloned in the pRZ5202 vector, exactly as described (Auble et al., 1986). Galactokinase assays were performed at 28 °C as described by McKenney et al. (1982) on cells harboring the pTA03 vector with the inserted promoters. The latter constructs were suitable for the stronger promoters, as they displayed little sensitivity of plasmid copy number [assayed as described (Auble et al., 1986)] to promoter strength. However, due to the relatively high background of the promoterless plasmid, pTA03 was less suitable for the assay of the weaker promoters. These were assayed in the pRZ5202 vector. All values are expressed as activity relative to the U5/6,D4/6 promoter [equivalent to the prm-up1 promoter; e.g., Auble et al. (1986); see Results], assayed under the same conditions in the same experiment; this ratio is referred to as RIV (for relative in vivo levels). Excellent agreement was found between the relative values determined in the two different assays for promoter variant U4/6,D6/6: this promoter was 2.4 ± 0.8 times as active as U5/6,D4/6 in the β-galactosidase assay and 2.6 ± 1.0 times as active in the galactokinase assay. Correction for differences in plasmid copy number were made, assuming a simple proportional relationship between copy number and activity. The largest corrections—those for U6/6,D6/6—amounted to a factor of 2 or less. Occasional series with a larger fluctuation were discarded.

RESULTS

The Promoters. The series of promoters that formed the basis for the investigations reported here is shown in Figure 1A. In our nomenclature we refer to the −10 and −35 regions as D (downstream) and U (upstream), respectively, and the number of base pairs that are consensus-like for each region are enumerated. The nine promoters are derivatives of the P_{RM} promoter of phage λ, which differs in two positions from the consensus sequence in both the −10 and −35 regions. At one of these four positions (−34) the consensus base pair is substantially more conserved (84%) than at the other three (44–54%) (Hawley & McClure, 1983). The base pairs in each region have been altered in a distinct order in two steps; the changes introduced in each region are independent of each other. As displayed, the series shows the "improvement" (in the sense of acquiring greater homology to the consensus sequence) of the −10 region in the context of three different −35 regions. In Figure 1B the promoters are arranged in a hierarchy determined by the number of base pairs that match the consensus −10 and −35 sequences.

In Vivo Studies. The chemically synthesized promoters were originally cloned into the vector pRZ5202 (Auble et al., 1986; see Materials and Methods) and then subcloned into the pKO1 derivative pTA03 (see Materials and Methods). In the former vector the promoters drive the expression of the gene for β-galactosidase and in the latter the expression of the galactokinase gene. Both gene products are easily assayed in crude cell extracts and thus provide a measure of the activity of the cloned promoters in vivo. The results of these assays are shown in Table I. The data have been corrected for variations in the plasmid copy number as detailed under Materials and Methods and then normalized to the expression of the U5/6,D4/6 variant to give the quantity defined as RIV. It can be seen that the activities span a range of nearly 3 orders of magnitude. The consensus promoter has the highest activity; it is 7 times more active than the wild-type *lac* promoter cloned in the same vector (data not shown). This indicates that the consensus promoter is among the strongest characterized thus far (Deuschle et al., 1986; Kammerer et al., 1986).

By comparison of the values of each group of three promoters shown in Table I, it is seen in Figure 2A that, for each of the −10 regions, the in vivo activities increase as the −35 region is converted to the consensus sequence in two successive steps. However, the relative increases in the in vivo expression are significantly larger for the worst −10 region (TAGATT) than for the best (TATAAT). A very similar trend is observed if promoters with the same −35 region are compared as a function of improving −10 regions.

In Vitro Studies. By use of the spectroscopic abortive initiation assay (Bertrand-Burggraf & Duane, 1984; see Materials and Methods) the kinetics of RNA polymerase binding to the promoters was also determined (McClure, 1980). In

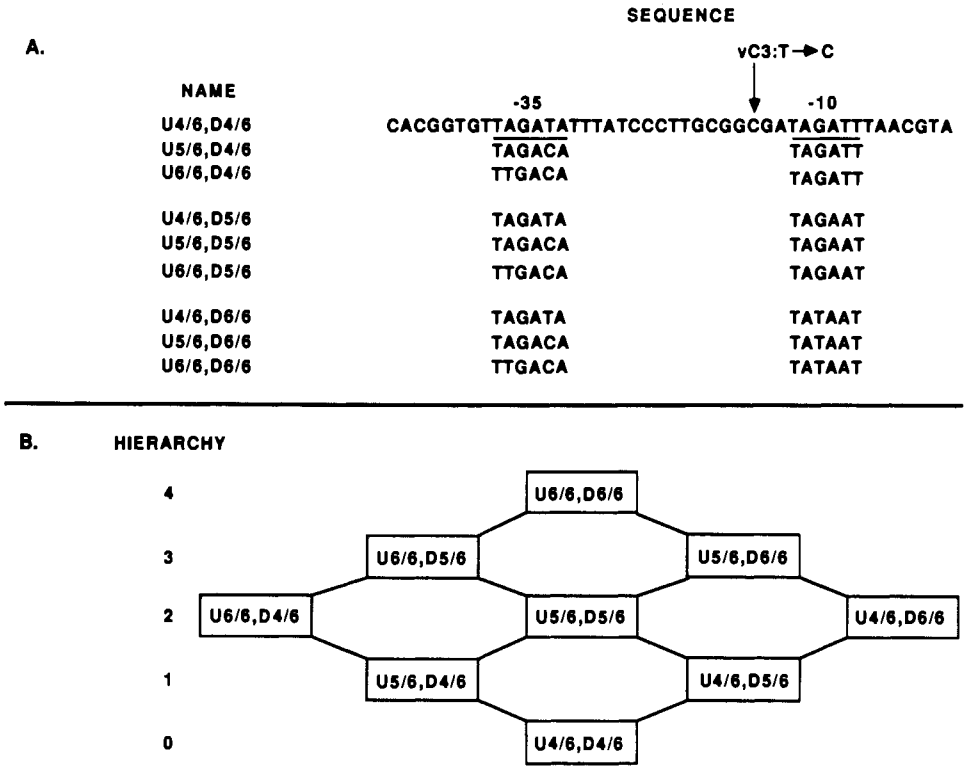

FIGURE 1: Promoters described in this work. (A) Partial sequence of the P_{RM} constructs investigated here. The location of the vC3 mutation is shown; this mutation is not relevant to the experiments presented in this paper. Each of the three groups of three variants has the same −10 region but a −35 region that becomes progressively more consensus-like from top to bottom. The −10 region of each group becomes more consensus-like in the same direction. (B) Arrangement of the variants by hierarchy. Each level is solely determined by the *number* of base pairs that differ from the consensus sequence, not their *location*. A left upward connector represents an alteration toward consensus in the −35 region; a right upward one, such a change in the −10 region.

Table I: Experimental Data for the Family of Nine Promoters

promoter	hierarchy[a]	homology score[b]	log RIV + 1[c]	log $K_B k_f$	log K_B	log k_f + 4[d]
U4/6,D4/6[e,f]	0	49.1	0.1 ± 0.1	2.3 ± 0.3	5.8 ± 0.6	0.5 ± 0.6
U5/6,D4/6[f]	1	52.6	1.0	4.1 ± 0.3	6.9 ± 0.6	1.2 ± 0.6
U6/6.D4/6	2	62.7	1.9 ± 0.2	4.9 ± 0.1	6.7 ± 0.5	2.2 ± 0.5
U4/6,D5/6[f]	1	52.1	1.2 ± 0.1	3.7 ± 0.4	(6.3)[g]	(1.4)[g]
U5/6,D5/6	2	55.6	1.k ± 0.1	5.3 ± 0.4	7.0 ± 0.5	2.3 ± 0.3
U6/6,D5/6	3	65.7	1.9 ± 0.1	7.0 ± 0.4	8.4 ± 0.5	2.6 ± 0.1
U4/6,D6/6	2	56.8	1.4 ± 0.1	5.3 ± 0.1	7.2 ± 0.4	2.1 ± 0.4
U5/6,D6/6	3	60.4	1.8 ± 0.04	7.5 ± 0.6	9.0 ± 0.6	2.5 ± 0.0
U6/6,D6/6	4	70.4	2.4 ± 0.2	(6.6)[h]	(9.8)[h]	2.8 ± 0.0

[a] Hierarchy level is determined by the degree of homology to the consensus sequences in the −10 and −35 regions. The fully consensus promoter is level 4, with the number decreasing to 1 for each mismatch. [b] Homology scores were calculated according to Mulligan et al. (1984). [c] This refers to the ratio of in vivo enzyme activities when driven by the promoter in question and when driven by U5/6,D4/6. Values are averages of two or more independent determinations, with standard deviations. Addition of 1 was so that algarithms would have a positive value. [d] Addition of 4 to the value was so that all would be positive. [e] Data from Hawley and McClure (1982). See text. [f] Measurements at 37 °C, converted to expected values at 25 °C (see text). The errors reflect the manipulation. [g] Estimated values obtained from the experimentally determined $K_B k_f$ and an interpolated value for k_f from a plot vs. the hierarchy levels. [h] Estimated values obtained from the experimentally determined k_f and an extrapolated value for $K_B k_f$ from Figure 6.

this assay formation of an open or functional complex is measured by the promoter-specific synthesis of a short RNA product. Our promoters direct the synthesis of UpApU from UpA and UTP. For each promoter a lag time, τ, corresponding to the average rate of formation of an open complex was determined as a function of RNA polymerase concentration. This allowed the process of open complex formation to be broken down into two steps. The first step is dependent on RNA polymerase concentration and is defined as the binding of RNA polymerase to the promoter. The second step is independent of RNA polymerase concentration and is referred to as the isomerization of the initial complex to give the open complex (McClure, 1980):

$$R + P \overset{\text{binding}}{\rightleftharpoons} RP_c \overset{\text{isomerization}}{\rightleftharpoons} RP_o$$

RNA polymerase (R) and promoter DNA (P) are assumed to be in rapid equilibrium (association constant K_B) with the inactive closed complex RP_c [the rapid equilibrium assumption has been verified for the parent promoter of the series, U4/6,D4/6 (P_{RM}), and also for U5/6,D4/6 (P_{RM}-up1) (Hawley & McClure, 1982; Auble et al., 1986)] which isomerizes in the rate-determining step (rate constant k_f) to the active open complex RP_o. Then a plot of τ vs. the reciprocal of the RNA polymerase concentration (a τ plot) is linear, with a slope $(K_B k_f)^{-1}$ and an intercept k_f^{-1}. It should be emphasized that the treatment of open complex formation as a two-step process is an oversimplification. However, the utility of this approach is that it allows the dissection of the kinetic data into two parameters (entirely analogous to the Michaelis–Menten treatment of enzymatic reactions), one (i.e., K_B) reflecting a

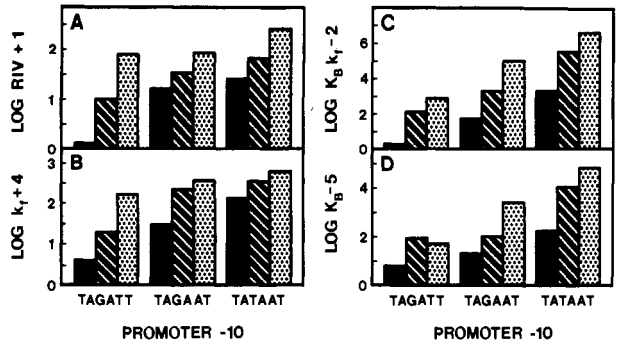

FIGURE 2: Pictorial representation of the data: In vivo promoter activities and kinetic parameters determined in vitro as a function of base changes in the −35 region, in the context of different −10 regions. The data are arranged in groups of three, each with a different −10 region. Within each group the −35 region becomes more consensus-like from left to right: solid bars, TAGATA; striped bars, TAGACA; dotted bars, TTGACA. Both measured and estimated (those in parentheses in Table I) values are included (see text). In panels A, B, C, and D, respectively, log RIV (the relative in vivo activities of promoter-driven β-galactosidase or galactokinase enzymes), log k_f, log K_B, and log $K_B k_f$ are plotted on the vertical axes.

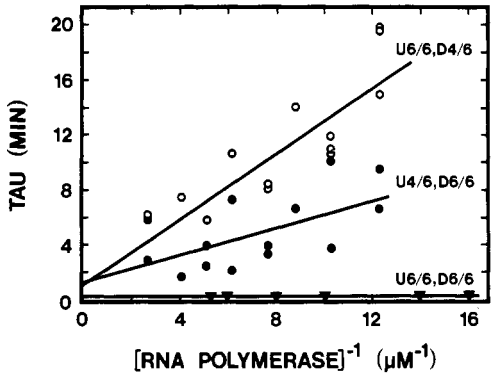

FIGURE 3: τ analysis of promoters D4/6,U6/6, D6/6,U4/6, and D6/6,U6/6. Lag times for the formation of open complexes, determined by the fluorescence assay at 25 °C, are plotted vs. the reciprocal of the RNA polymerase concentration (corrected for the fraction of active enzyme. Note the much smaller y intercept (corresponding to a faster rate of isomerization) and shallower slope (corresponding to a faster overall reaction rate $K_B k_f$) for the promoter with consensus −10 and −35 sequences (D6/6,U6/6).

bimolecular process involving free DNA and RNA polymerase and the other (k_f) the subsequent unimolecular reactions. For the P_R promoter of phage λ, Roe et al. (1985) found good agreement between the value of the parameter K_B and the experimentally determined binding constant for RNA polymerase–promoter interaction at temperatures below 10 °C, where closed complex formation should be favored.

In an attempt to slow the rates of open complex formation for the strongest promoters to an experimentally accessible range, the measurements were carried out at 25 °C. As the range of reaction rates was greater than our experimental window, meaningful data at this temperature could only be obtained for the six promoters in the top three hierarchical levels. τ plots for three promoters with consensus −10 or −35 regions or both are shown in Figure 3. These plots are representative of the range of rates of open complex formation we observed at 25 °C and of the various extents of scatter we see in our τ plots.

Computer fits to the data allowed the determination of the kinetic parameters characteristic of each promoter as well as estimates of their associated errors. In Table I the logarithms of the values are given, as these were used in subsequent analyses. The largest errors were in K_B (an average of 0.5 log unit), compared to an average of 0.3 log unit in $K_B k_f$ and k_f. This is not surprising, as the latter two parameters are determined directly from the slopes and intercepts of the τ plots while K_B is determined from their ratio. The errors are quite large, but meaningful trends and correlations can still be obtained from the data since the values of k_f, K_B, and $K_B k_f$ span a strikingly wide range.

The lag times for the binding of RNA polymerase to the two promoters in level 1, and to U6/6,D4/6 as a reference, were determined at 37 °C. The data obtained for U5/6,D4/6 (P_{RM}-up1) allow a comparison with previously determined values for this promoter, by use of a chromatographic method for the detection of the abortive product UpApu (Hawley & McClure, 1982; Auble et al., 1986). The values for k_f are in good agreement, but the K_B's determined here are a factor of 3 lower than those reported earlier, a relatively small discrepancy. To complete the set, we used the relative values for the kinetic parameters between U5/6,D4/6 and U4/6,D4/6 (P_{RM}) determined by Hawley and McClure (1982), also at 37 °C. [Their promoters differ from the ones used here in that they do not contain the vC3 mutation (see Figure 1); this mutation does not appear to affect promoter function as determined from a comparison of our data (Auble et al., 1986) and those of Hawley and McClure (1982)]. The values obtained at 37 °C have been adjusted to reflect the situation at 25 °C by using the temperature dependences determined for the U6/6,D4/6 promoter (at the higher temperature K_B is higher by a factor of 3, k_f by a factor of 2.3), and all data are presented in Table I. In combining the data determined at different temperatures in this fashion, we make the implicit assumption that the temperature dependence of all kinetic parameters for the other promoters measured at 37 °C is identical with that for U6/6,D4/6. While this assumption seems reasonable, it remains otherwise unsubstantiated; the fact that the corrections made are relatively small would tend to minimize any differences in the behavior of the various promoters [it should be pointed out that in a study of the interaction of *lac* repressor with mutant operators a DNA sequence dependence of the effect of temperature has been demonstrated by Mossing and Record (1985)].

The data set is incomplete for two promoters; the bracketed values are given for completeness only and have not been used in quantitative analyses. For the consensus promoter (U6/6,D6/6), the τ plot is nearly flat; we have estimated a lower limit for K_B of 7×10^8 M^{-1}, on the basis of the shallowest slope we can measure, given the experimental error of the data. To get a better approximation, we have estimated the value of $K_B k_f$ by extrapolation of Figure 5 (below). For variant U4/6,D5/6 the τ plot has an uninterpretable negative intercept, making the evaluation of k_f from the plot impossible even though the product $K_B k_f$ is known with reasonable accuracy from the slope. In this case, k_f was obtained by interpolation of a plot of log k_f vs. promoter hierarchy (not shown).

From Table I it is seen that the promoter with consensus −10 and −35 sequences apparently functions optimally in vitro as well: the fastest rates of open complex formation are observed with U6/6,D6/6. In the next several figures we have plotted the data in Table I in various ways, to illustrate aspects of the role of the −10 and −35 regions in determining promoter function. From Figure 2B it is apparent that for each −10 sequence, the values of k_f increase as the −35 region is improved, with the greatest proportional changes occurring for the poorest −10. In contrast, the greatest jump in K_B values occurs for the consensus −10 (Figure 2D); these opposing trends cancel in the product $K_B k_f$, which exhibits similar

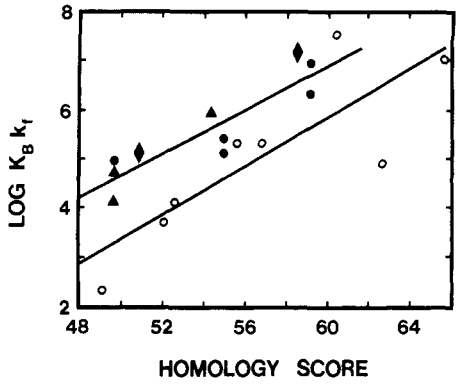

FIGURE 4: Relationship between the logarithms of the slopes of the τ plots (i.e., log $K_B k_f$) and the homology scores for the series of promoters studied here and selected other promoters from Mulligan et al. (1983). (○) Data for $K_B k_f$ from Table I; homology scores have been calculated as described by Mulligan et al. (1983). The slope of the least-squares fit to the data is 0.25, and the correlation coefficient is 0.83. Filled symbols: Data compiled by McClure et al. (1983); only promoters and their variants involving modifications in the −10 and −35 hexamers have been included; the slope is 0.22, and the correlation coefficient is 0.90. (●) *lac* promoter variants; (▲) P_{RM} promoter variants; (♦) P_R promoter variants.

relative increases for all three −10 regions (Figure 2C). For all three variables the situation is very similar to the above if, instead, the −10 region is improved in the context of each of the three −35 regions (plots not shown).

Comparison of in Vivo and in Vitro Results. It has been argued that it is the overall rate of promoter binding—the product $K_B k_f$—that correlates best with in vivo activity (McClure, 1985). On the other hand, our prior studies with spacer variants of the U5/6,D4/6 promoter were best interpreted by correlating in vivo activities with k_f (Auble et al., 1986). We have plotted the in vivo results vs. the three kinetic parameters of Table I and used a linear least-squares fit to determine correlations. The best correlation (0.96) is obtained between log RIV and log k_f; correlation coefficients of 0.85 and 0.70 are obtained from fits of log RIV vs. log $K_B k_f$ and log K_B, respectively (plots not shown). Statistical comparison of these fits (using Fisher's Z transformation) indicates however that due to the small number of points on each plot (seven or eight) the differences in correlation coefficients are not significant. Inclusion of the estimated values for promoters U4/6,D5/6 and U6/6,D4/6 improves all fits without changing their relative ranking, but again the differeces are not statistically significant. It is thus not possible to deduce, from our data, how the kinetic parameters determined in vitro relate to promoter utilization in vivo. A similar conclusion is reached when the data compiled by McClure (1983) are analyzed in the above fashion (not shown).

Correlation of in Vitro Kinetics with Homology Scores. McClure and co-workers (Mulligan et al., 1984) quantified the homology of promoters to the sequence of a consensus promoter by defining a homology score. The contribution of various conserved features (the −10 and −35 regions; regions up- and downstream of these; the length of the spacer separating the −10 and −35 regions) to the homology score is weighted according to the extent of conservation of the feature. A plot of log $K_B k_f$ vs. the homology score for 31 promoters gave a correlation coefficient of 0.83, and thus this represents the first successful attempt at predicting promoter activity from promoter sequence. We have calculated homology scores of our collection of promoters (see Table I) and found a reasonable correlation (correlation coefficient = 0.83) with log $K_B k_f$ as well (see Figure 4). However, the slope of our plot was twice as steep as the slope of the published plot of

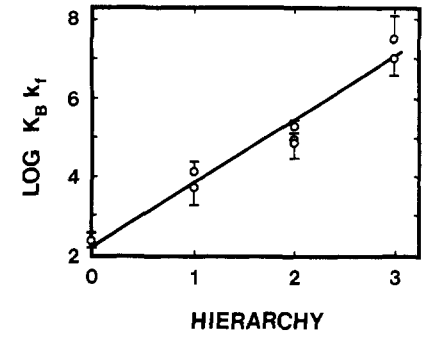

FIGURE 5: Correlation between the experimentally determined log $K_B k_f$ of the various promoters and their hierarchy levels. Values are from Table I. The correlation coefficient of the linear least-squares fit to the data is 0.98.

Mulligan et al. (1984) (0.25 vs. 0.11), indicating that our data are more sensitive to the homology score than the aggregate of the data used by Mulligan et al. (1984). If their data set is limited to include only families of promoters differing at their −10 and −35 regions, a slope of 0.22 is obtained (and a correlation coefficient of 0.90) (see Figure 4), in good agreement with out results. The lower $K_B k_f$ values characteristic of our values are due in large part to the difference in temperature at which the two sets of data were collected. We thus conclude that the current version of the homology score does not weight the −10 and −35 regions heavily enough, as compared to other regions. This is another indication of the overwhelming importance of the sequences at −10 and −35 for the activity of *E. coli* promoters.

An alternative representation of our data is a plot of log $K_B k_f$ vs. the hierarchy level (defined in Figure 1): An excellent correlation (correlation coefficient 0.98) is obtained (see Figure 5). Due to the small number of points, this correlation is statistically only marginally better than that obtained in the plot vs. homology score, above ($p < 0.1$). While it was not expected that a linear relationship would exist between log $K_B k_f$ and promoter hierarchy, the fact that this relationship occurs over such a wide range of promoter strengths has several implications for the effect of a base change in the −10 or −35 region. These will be considered under Discussion.

DISCUSSION

Influence of Promoter Regions Other Than −10 and −35. For the series of promoters investigated here, clearly the consensus promoter is best both in vivo and in vitro. However other regions of a promoter can also play an important role in its activity, notably by affecting processes that occur subsequent to open complex formation (Deuschle et al., 1986; Kammerer et al., 1986). It was found that the region downstream of the start site can have a drastic effect on the in vivo expression from a promoter, for example, leading to relatively low activity of a promoter with consensus −10 and −35 regions. By investigating the family of promoters shown in Figure 1, which differ from each other only in their −10 and −35 regions, we have circumvented this potential complication.

Comparison of Mutations in the −10 and −35 Regions. In agreement with other studies [e.g., Youderian et al. (1982); other references can be found in von Hippel et al. (1984) and McClure (1985)], we find that mutations that increase the homology of a promoter with the consensus promoter sequence lead to enhanced promoter function. Our focus here is on the comparison of the effects of base pair changes in the −10 and −35 regions. Initial characterizations of promoter variants (Shih & Gussin, 1983; Hawley & McClure, 1982, 1981) were interpreted as indicating that mutations in the −10 region

affected mostly k_f, while those in the −35 region affected K_B. Our finding that k_f is affected by alterations in both the −10 and the −35 regions (e.g., see Figure 2) is contrary to this expectation. Further inspection of the data in Table I indicates that k_f, K_B, and K_Bk_f each appear to be affected similarly by mutations in either region: The values are determined primarily by the promoter hierarchy and not by the region of the promoter in which the change is introduced. The similarity in the values for promoters of hierarchy level 2 is especially striking, as this class is composed of promoters that have two nonconsensus bases in either the −10 region or the −35 region (U6/6,D4/6 and U4/6,D6/6), as well as a promoter that has one mutation in each region (U5/6,D5/6). Thus, our in vitro results demonstrate that base changes in the −10 and −35 regions of promoters have similar effects on the kinetics of open complex formation at the promoters regardless of sequence conservation (see below), suggesting that the positions investigated are functionally equivalent. This would support a model for which simultaneous binding of RNA polymerase to both the −10 and −35 regions [e.g., Stefano and Gralla (1982a)] was necessary for at least the rate-limiting step toward formation of an open complex.

The situation is much the same for the in vivo data, but an exception is readily apparent: The U6/6,D4/6 promoter is significantly more active in vivo than the other two promoters of the same hierarchy; we have verified that this high level is not peculiarity of the particular U6/6,D4/6 clone used (data not shown). An obvious difference between the in vivo and the in vitro data is that the former are obtained with promoters on supercoiled plasmid DNAs inside the cell, while the latter are for promoters on linearized plasmid DNA. A differential activation of the U6/6,D4/6 promoter by supercoiling would explain the above observation.

Sequence Conservation and Functional Significance. The underlying assumption of statistical treatments with the goal of correlating promoter sequence and promoter function has been that changing a base pair that is strongly conserved among promoters should have a larger effect than the change of a less conserved base pair (Mulligan et al., 1984; Berg & von Hippel, 1987). However, the data presented here indicate that the two mutations toward consensus introduced into U4/6,D4/6 in each region have qualitatively and quantitatively equivalent effects, even though the extent of sequence conservation differs substantially for the two base pairs affected. This inference is also drawn from the comparison of promoters belonging to the second hierarchical class: the promoters with double base pair changes in either the −10 region or the −35 region (U6/6,D4/6 and U4/6,D6/6) behave similarly to the promoter with single changes in both regions (U5/6,D5/6). The simplest interpretation (necessarily tentative however because only one of the two changes was introduced as a single alteration) of this observation is that the effects of the changes are equivalent. Yet the substitution at −34, characteristic of the U5/6 → U6/6 change, is at a much more highly conserved base pair [a T in the top strand for 84% of the promoters of the Hawley and McClure (1983) collection]. [It has recently been shown that RNA polymerase contacts the 5-methyl group of this base pair in the major groove (Dubendorff et al., 1987)]. Apparently, sequence conservation and functional significance are not correlated for these base pairs; a similar observation was also reported for the most downstream position of the −35 region (Rossi et al., 1983).

The above conclusion is further substantiated by the excellent fit obtained between log K_Bk_f and promoter hierarchy (see Figure 5), despite the fact that no allowances have been made for the different extents of sequence conservation. In fact, the correlation is at least as good as that between log K_Bk_f and the homology score (Figure 4), which attempts to take differences in sequence conservation into account. The lack of quantitative correlation between sequence conservation and functional importance leads in the homology score to an overestimation of the effect of a change at the −34 base pair and is also the reason for the underestimation of the contributions of the −10 and −35 regions (Mulligan et al., 1983) (see Results). It should be pointed out however that a plot like that of Figure 5 can only be constructed for a family of promoters such as studied in this paper. The homology score (Mulligan et al., 1983) is a much more generally applicable indicator of promoter function despite its limitations.

Sequence Context Effects. Base pair changes in the −10 and −35 regions have been found to be equivalent regardless of which of the two regions is concerned or the degree of sequence conservation at the altered position. We now address the question whether the effect of a base change is also independent of the particular promoter sequence into which it is introduced. The excellent fit between hierarchy level and log K_Bk_f (Figure 5) demonstrates that this independence indeed holds for the parameter K_Bk_f. This appears to be due to compensatory trends in k_f and K_B: The effect on k_f of a particular base change toward consensus is greatest, as the promoter into which this base change is introduced differs in more positions from the consensus −10 and −35 regions (see Figure 2B). K_B shows the opposite trend (Figure 2D). The results obtained here for K_Bk_f may be of general significance: Stefano and Gralla (1982b) reported a sequence context independent behavior of the overall rate of promoter binding (equivalent to our K_Bk_f), while investigating base changes in the lactose promoter at positions both similar to and different from those studied here. It is possible that the greater sensitivity of the k_f parameter of U4/6,D4/6 (the P_{RM} promoter) to sequence changes in the promoter reflects a physiologically significant enhanced sensitivity of its k_f to another stimulus as well: The promoter activated by cI protein has an increased k_f (see below).

Possible Kinetic Determinants of Promoter Activity in Vivo. Our experimental results do not allow an unambiguous assignment of a particular kinetic parameter as the determinant of promoter activity in vivo: While the calculated correlation coefficient of the plot of log RIV vs. log k_f is greater than that for plots of log RIV vs. log K_Bk_f or log K_B, the difference is not statistically significant (see Results). Two lines of evidence point to the importance of k_f in this context and eliminate K_B as a potential sole determinant of promoter activity in vivo. Auble et al. (1986) investigated a series of spacer variants of U5/6,D4/6, for which K_Bk_f had similar values, but K_B and k_f showed changes in opposing directions. The in vivo activities of these promoters followed the trends in k_f. The reports (Hawley & McClure, 1982; Shih & Gussin, 1983) that cI protein activates transcription initiation from the P_{RM} promoter by increasing k_f are also consistent with this parameter being rate determining in vivo. A definitive test to ascertain whether only k_f determines in vivo promoter function (in which case promoter *binding* would not be limiting in vivo), or whether the product K_Bk_f does so, would be provided by studying an as yet unavailable promoter variant for which K_B, but not k_f, is altered.

ACKNOWLEDGMENTS

We thank P. Martello, G. Brooke, and K. Boothby for help in the construction and sequencing of promoter variants, Dr. McClish for help with statistical analysis, and Drs. M. T.

Record, Jr., and C. Harrison and our laboratory colleagues for their comments on a version of the manuscript.

REFERENCES

Auble, D. T., Allen, T. L., & deHaseth, P. L. (1986) *J. Biol. Chem. 261*, 11202–11206.

Berg, O. G., & von Hippel, P. H. (1987) *J. Mol. Biol. 193*, 723–750.

Bertrand-Burggraf, E., & Daune, M. (1984) *Nucleic Acids Res. 12*, 1697–1706.

Brosius, J., Erfle, M., & Storella, J. (1985) *J. Biol. Chem. 260*, 3539–3541.

Bruzik, J. P., Auble, D. T., & deHaseth, P. L. (1987) *Biochemistry 26*, 950–956.

Chamberlin, M. J. (1974) *Annu. Rev. Biochem. 43*, 721–775.

Deuschle, U., Kammerer, W., Gentz, R., & Bujard, H. (1986) *EMBO J. 5*, 2987–2994.

Dubendorff, J. W., deHaseth, P. L., Rosendahl, M. S., & Caruthers, M. H. (1987) *J. Biol. Chem. 262*, 892–898.

Hawley, D. K., & McClure, W. R. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6381–6385.

Hawley, D. K., & McClure, W. R. (1982) *J. Mol. Biol. 157*, 493–525.

Hawley, D. K., & McClure, W. R. (1983) *Nucleic Acids Res. 11*, 2237–2255.

Inouye, S., & Inouye, M. (1985) *Nucleic Acids Res. 13*, 263–267.

Kammerer, W., Deuschle, U., Gentz, R., & Bujard, H. (1986) *EMBO J. 5*, 2995–3000.

Mandecki, W., Goldman, R. A., Powell, B. S., & Caruthers, M. H. (1985) *J. Bacteriol. 164*, 1353–1355.

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

Maquat, L. E., & Reznikoff, W. R. (1978) *J. Mol. Biol. 125*, 467–490.

McClure, W. R. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5634–5638.

McClure, W. R. (1983) in *Biochemistry of Metabolic Processes* (Lennon, D. L. F., Stratman, F. W., & Zahlten, R. N., Eds.) pp 207–217, Elsevier, New York.

McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171–204.

McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C., & Rosenberg, M. (1982) in *Gene Amplification and Analysis* (Chirikjian, J. G., & Papas, T., Eds.) pp 734–739, Elsevier/North-Holland, New York.

Mossing, M. C., & Record, M. T., Jr. (1985) *J. Mol. Biol. 186*, 295–306.

Mulligan, M. E., Hawley, D. K., Entriken, R., & McClure, W. R. (1984) *Nucleic Acids Res. 12*, 789–800.

Mulligan, M. E., Brosius, J., & McClure, W. R. (1985) *J. Biol. Chem. 260*, 3529–3538.

Munson, L. (1983) Thesis, University of Wisconsin, Madison.

Roe, J.-H., Burgess, R. R., & Record, M. T. (1985) *J. Mol. Biol. 184*, 441–453.

Rossi, J. J., Soberon, X., Marumoto, Y., McMahon, J., & Itakura, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3203–3207.

Shih, M.-C., & Gussin, G. N. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 496–500.

Stefano, J. E., & Gralla, J. D. (1982a) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1069–1072.

Stefano, J. E., & Gralla, J. D. (1982b) *J. Biol. Chem. 257*, 13924–13929.

Stefano, J. E., Ackerson, J. W., & Gralla, J. C. (1980) *Nucleic Acids Res. 8*, 2709–2723.

von Hippel, P. H., Baer, D. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.

Yarbrough, L. R., Schlageck, J. G., & Baughman, M. (1979) *J. Biol. Chem. 254*, 12069–12073.

Youderian, P., Bouvier, S., & Suskind, M. M. (1982) *Cell (Cambridge, Mass.) 30*, 843–853.