Escherichia coli promoter –10 and –35 region homologies correlate with binding and isomerization kinetics

Gary M. STUDNICKA

Ingene Inc., 1545 Seventeenth Street, Santa Monica, CA 90404-9954, U.S.A.

The DNA promoter sequence at which gene transcription is initiated in *Escherichia coli* contains two distinct regions of conserved nucleotides. Coefficients of homology to the –10 region and the –35 region were computed for many different prokaryotic promoters. Linear equations were derived that relate the degree of homology for each promoter region to the two kinetic constants that describe the interaction of RNA polymerase with the promoter site on DNA: the strength K_B of binding to form closed complex, and the rate k_t of isomerization to form open complex. A graphical plot of –35 versus –10 promoter region homologies for many promoters suggests that certain classes of prokaryotic operons might utilize differential degrees of consensus homology in these two regions to achieve specific control over k_t and K_B , in addition to modulating overall promoter strength.

INTRODUCTION

The initiation of transcription at promoter sites is a major control point for the regulation of gene expression. Pribnow (1975) found that specific conserved sequences for RNA polymerase binding at promoter sites were located about ten bases upstream from the transcription initiation points on DNA. Gilbert (1976) showed that a separate conserved mutationally defined ‘recognition site’ existed, and that it was located about 16–18 bases upstream from the Pribnow sequence, near position –35. This observation eventually led to a model for promoter function, termed the ‘bipartite model’ (McClure, 1982), in which the –35 region determines the strength K_B of formation of closed complex (sequence-specific binding), and the –10 region determines the rate k_t of isomerization to open complex (helix ‘melting’). Johnston & McClure (1976) developed the ‘abortive initiation assay’ to measure the kinetic constants associated with transcription, and to separate the binding constant from the isomerization constant. This was a major step forward, allowing for the first time a direct investigation of the molecular mechanism of initiation. Since then, many elegant kinetic experiments have revealed details of the initiation complex for a variety of promoters. Mulligan *et al.* (1984) showed that the statistical matching coefficient for overall promoter homology is linearly related to the overall kinetic constant $K_B k_t$ for formation of open complex, but at that time they did not detect a clear relationship between the –35 homology and K_B or between the –10 homology and k_t . To test the roles played by the –35 and –10 regions in the bipartite model, and to determine whether promoters of *Escherichia coli* (and its plasmids and bacteriophages) employ strength or weakness in either the –35 or the –10 region for specific regulatory purposes, I used a computer program (Studnicka, 1986) to calculate coefficients of homology for many known prokaryotic promoters. Some of them were compared with previously published kinetic constants, and two equations were derived to relate DNA homologies to initiation kinetics. This study also suggests that strength in either the –35 region or the –10 region might be

employed according to a consistent pattern by specific classes of genes to determine the slow step of initiation.

METHOD

The computer program, described in detail previously (Studnicka, 1986), uses consensus matrices and statistical weighting to compute a coefficient of matching for the –10 region and for the –35 region separately, and it indicates the significance of a match by how many standard deviation units (‘sigmas’) that homology is from the mean homology obtained for a ‘control’ sequence of random DNA. This makes it easy to estimate how often an equal or better homology would be found purely by chance. To create a consensus matrix, DNA sequences from 112 prokaryotic promoters were aligned (Hawley & McClure, 1983) to maximize consensus homology. Associated with each consensus position are four numbers, showing how often A, C, G or T appeared there in the group of known promoters. When this consensus matrix is aligned with a very long fragment of random DNA (equimolar in the four bases), and when coefficients of matching are computed for every possible alignment position, the mean matching coefficient is defined to be 0.0 sigmas. All of the coefficients are normally distributed around that mean, forming a typical bell-shaped curve (Bevington, 1969, pp. 43–49), whose standard deviation is defined to be exactly 1.0 sigma. Since the distribution is symmetrical about the mean, half the matching coefficients expressed in sigma units (‘normalized’ distances from the mean) are negative and half are positive. Integration under the normal curve shows (Bevington, 1969, pp. 304–309) that 68% of the homology coefficients lie between –1 and +1 sigmas, that 95% lie between –2 and +2 sigmas, and that 99.7% lie between –3 and +3 sigmas. Thus a homology whose matching coefficient is greater than +2 sigmas can be expected to occur by chance once in about 50 alignment positions, a match better than +3 sigmas occurs once in 1000, and better than +4 occurs once in 30000 alignments. The computer program calculates a matching coefficient (expressed in sigmas) for the –35

region, and a separate one for the -10 region, for each promoter under investigation. Spacing between the two regions does not influence either of the two coefficients in this analysis. Using the nomenclature (Nomenclature Committee of the International Union of Biochemistry, 1985) for incompletely specified nucleic acid sequences (e.g. $r = A/G$, $y = C/T$, $w = A/T$, $d = A/G/T$, $h = A/C/T$), the consensus matrix for the -35 promoter region is 'th...awwha.ttcTTGAcwt..t.' and '...wt.tg.TAtAaTd..' describes the -10 consensus matrix (Studnicka, 1986). In each consensus upper-case letters indicate strongly conserved bases, lower-case letters are for moderately conserved, and periods denote non-conserved nucleotide positions. Thus more than merely the two conserved hexamers (TTGAcw and TAtAaT) are taken into account by the computer program, although the influence of weakly conserved positions is appropriately smaller.

RESULTS AND DISCUSSION

Calculation of homology coefficients

Homology coefficients were calculated for many promoters (Hawley & McClure, 1983; Fournier & Ozeki, 1985; Gentz & Bujard, 1985; Mulligan *et al.*, 1985; Cowing *et al.*, 1985; Lindahl & Zengel, 1986), and each promoter was plotted on the graph in Fig. 1(a). Four weak promoters (*malT*, *groE*, *fdI* and *dnaKp2*) are slightly off the graph. The names of some *E. coli* genes were modified to reflect the currently recommended genetic nomenclature (Bachmann, 1983). Homology to the -35 region increases along the horizontal axis, whereas -10 homology increases along the vertical axis, so that overall promoter homology increases from lower left to upper right. Perpendicular to this diagonal, the relative contributions of the -35 and -10 promoter regions can change without greatly affecting the overall homology. For example, λpR and *Tn10pOUT* are promoters with similar overall homology and similar overall kinetic activity (Simons *et al.*, 1983), but λpR has a prominent -35 region whereas *Tn10pOUT* has a prominent -10 region instead. Some of the promoters (e.g. *metT*) exhibit simple recognition patterns, in which the -35 and the -10 regions are both about equally strong, and give rise to a single strong promoter homology with only one major alignment frame. Others (e.g. *lpp*) consist of multiple overlapping weak homologies (details not shown), in which the RNA polymerase might bind in many alternative ways, each with different positions and spacings. For the promoters in Fig. 1(a), the -35 homologies vary over a large range (1 to 5 sigmas, with a mean of 2.9 ± 1.1 sigmas) versus a narrower interval (2 to 4 sigmas, with a mean of 3.1 ± 0.7) for the -10 homologies. The most homologous promoters overall (e.g. *recA*, *G4D*, *rpsL*, *tac17*, *T522*)

have combined overall matching coefficients (considering the -35 and -10 regions together) of about 6 sigmas (they are expected to occur by chance alone very rarely, once in a few million bases), whereas the weakest ones (e.g. *malP*, *rpoB*) are only about 2 sigmas (expected to occur very frequently, every 50–100 bases). The range of statistically expected promoter homology frequencies agrees well with the observed dynamic range $K_B k_t$ in promoter activity of about 10^4 (McClure, 1985). Although the exact value of each homology coefficient will vary slightly with the choice of sequences that make up the consensus table and with the particular method of scoring homology, the overall organization of Fig. 1(a) should remain relatively invariant with the use of any reasonable homology scheme, so the results do not depend on using this particular computer program to calculate the homology coefficients.

Correlation with the kinetics of polymerase binding and isomerization

The 'abortive initiation assay' for prokaryotic promoter activity, developed by William McClure, is a promoter-specific steady-state reaction that has been of great value, permitting investigators to separate the contributions of binding and isomerization in the transcription initiation complex, and to measure each one accurately. When all of the components for transcription *in vitro* are provided, except that at least one RNA triphosphate is omitted, RNA polymerase binds to the promoter, isomerizes to open complex, and repeatedly synthesizes a very short oligoribonucleotide without dissociating from the catalytic complex between steps. The assay takes advantage of the fact that a temporal lag exists between the time the DNA, polymerase and triphosphates are mixed together and the time at which steady-state synthesis of the specific radioactively labelled oligonucleotide is reached by the RNA polymerase molecules trapped in open complexes at the promoter sites. Lag time ('tau') is plotted as a function of inverse polymerase concentration for each experiment, and McClure (1980) has shown that lag time at 'infinite' polymerase concentration (the vertical intercept, $1/k_t$) determines the isomerization constant k_t (measured in s^{-1}) and the molar dependence of lag time upon polymerase concentration (the slope of the line, $1/K_B k_t$) determines the binding constant K_B (measured in M^{-1}). From the 'tau plots' for each promoter a 'promoter selectivity map' can be constructed, such as Fig. 1(b), in which the vertical axis represents the logarithm of the isomerization constant k_t and the horizontal axis is the logarithm of the binding constant K_B . The data points for this map were compiled from previously published work (McClure, 1980; Hawley & McClure, 1980; McClure, 1982; Simons *et al.*, 1983; Dayton *et al.*, 1984; Mulligan *et al.*, 1985; Herbert *et al.*, 1986; Yu &

Fig. 1. Correspondence between promoter-region homology and initiation kinetics

(a) Consensus homology of 178 promoters. The vertical axis represents the homology (in sigma units) of the -10 region alone, and the horizontal axis is the homology of the -35 region alone. The number in brackets below the name of each promoter designates how many 'spacer' nucleotides are between its two hexameric regions. Where different spacings yield almost equally favourable homologies in the computer analysis, the two best spacings are shown. (b) Binding and isomerization kinetics of 19 promoters. The vertical axis is the logarithm of k_t (the isomerization constant, in s^{-1}), and the horizontal axis is the logarithm of K_B (the binding constant, in M^{-1}), obtained from the 'tau' plot (McClure, 1980). Each X symbol shows the kinetic constants predicted from the sequence homology in (a), by using the mapping functions. In both (a) and (b) promoters enclosed in ovals or rectangles were used to compute the mapping functions.

Reznikoff, 1986). If the bipartite model accurately represents promoter kinetics, then k_t should be correlated with the degree of -10 region homology and K_B with -35 region homology. The present study shows that a mapping function exists between the promoter homology map in Fig. 1(a) and the promoter selectivity map in Fig. 1(b), but that the relationship is not as simple as predicted by the bipartite model. As a first approximation, this correlation can be described using two linear equations:

$$\begin{cases} \log k_t = C_t + T_{t,-10}H_{-10} + T_{t,-35}H_{-35} \\ \log K_B = C_B + T_{B,-10}H_{-10} + T_{B,-35}H_{-35} \end{cases} \quad (1)$$

The pair of constants (C_B, C_t) linearly translates the origin (0,0) of the homology graph in Fig. 1(a) to the point (C_B, C_t) of the kinetics graph in Fig. 1(b), and the four values $T_{t,-10}$, $T_{t,-35}$, $T_{B,-10}$ and $T_{B,-35}$ together form a matrix T that linearly transforms the (H_{-35}, H_{-10}) homology co-ordinates into the ($\log K_B, \log k_t$) kinetic co-ordinates. I have obtained estimates of all six coefficients by performing two independent multiple linear regressions of the general form shown below, where (x, y) is the (H_{-35}, H_{-10}) homology co-ordinate and (x', y') is the ($\log K_B, \log k_t$) kinetic co-ordinate:

$$\begin{cases} y' = A + By + Cx \\ x' = D + Ey + Fx \end{cases} \quad (2)$$

Eight 17-spaced promoters (λpRM , $\lambda pRx3$, $\lambda pRMup1$, $Tn10pIN$, λpR , $Tn10pOUT$, T7A3 and *tac17*), enclosed by ovals in Figs. 1(a) and 1(b), were chosen to compute the mapping coefficients, which yield these two linear equations:

$$\begin{cases} \log k_t = -5.8 + 1.0H_{-10} + 0.2H_{-35} \\ \log K_B = +5.5 + 0.3H_{-10} + 0.6H_{-35} \end{cases} \quad (3)$$

Since total promoter activity is $K_B k_t$, it follows that:

$$\log K_B k_t = \log K_B + \log k_t = -0.3 + 1.3H_{-10} + 0.8H_{-35} \quad (4)$$

The inverse T^{-1} of the matrix T was also computed, giving two more equations, which reverse-transform from kinetic co-ordinates back to homology co-ordinates:

$$\begin{cases} H_{-10} = 8.7 + 1.1\log k_t - 0.4\log K_B \\ H_{-35} = -13.8 - 0.6\log k_t + 1.9\log K_B \end{cases} \quad (5)$$

The *lacUV5*, *lacZp1*, *galEp2*, *tac18* and T7A2 promoters (which are shown enclosed in rectangles) are all greatly shifted to the right relative to the other promoters on the homology graph in Fig. 1(a), as compared with their positions on the kinetics graph in Fig. 1(b). Their behaviour seems to be anomalous until it is realized that they are all 18-spaced promoters. Since they behave differently from the 17-spaced ones, a similar but independent set of equations was obtained for these five 18-spaced promoters:

$$\begin{cases} \log k_t = -5.7 + 1.3H_{-10} - 0.1H_{-35} \\ \log K_B = +4.5 - 0.2H_{-10} + 1.0H_{-35} \end{cases} \quad (6)$$

$$\log K_B k_t = \log K_B + \log k_t = -1.2 + 1.1H_{-10} + 0.9H_{-35} \quad (7)$$

$$\begin{cases} H_{-10} = 4.3 + 0.8\log k_t + 0.1\log K_B \\ H_{-35} = -3.6 + 0.2\log k_t + 1.0\log K_B \end{cases} \quad (8)$$

Graphical interpretation of the mapping function

Fig. 2(b) represents the entire promoter selectivity map from Fig. 1(b), and it shows how the T matrices for 17-

and 18-spaced promoters (eqns. 3 and 6) linearly transform a unit square from the 'other' ($-35, -10$) co-ordinate system into 'this' (bind, isom) system. The black square in the centre (labelled 'unit') has sides measuring exactly 1 log unit along each of the axes in the (bind, isom) co-ordinate system, and it is centred on the point (8.0, -2.0). There is a similar unit square in Fig. 2(a), which has sides measuring exactly 1 sigma unit along each of the axes in its own ($-35, -10$) system, and which is centred on the point (3.5, 3.5). The lightly shaded parallelogram labelled '[17]' in Fig. 2(b) shows how the black unit square from Fig. 2(a) is projected on to Fig. 2(b) by the mapping function (eqn. 3) for 17-spaced promoters. The arrow labelled ' -10 ' on this parallelogram indicates that movement of 1 sigma along the -10 axis of the unit square in Fig. 2(a) corresponds to an increase of 1.0 in $\log k_t$ and 0.3 in $\log K_B$. The ' -35 ' arrow shows that movement of 1 sigma along the -35 axis in Fig. 2(a) yields an increase of 0.2 in $\log k_t$ and 0.6 in $\log K_B$. This suggests that, for 17-spaced promoters, an up-mutation in the -10 region should greatly increase k_t and slight increase K_B , whereas an up-mutation in the -35 region should greatly increase K_B and slightly increase k_t . Examples of two such -35 region mutations are λpRM going to $\lambda pRMup1$ and $\lambda pRx3$ going to λpR (McClure, 1982). Since the -10 side of the parallelogram is almost twice as long as the -35 side, a 1-sigma up-mutation in the -10 region should increase the overall promoter activity $K_B k_t$ by about twice as much as an equivalent 1-sigma up-mutation in the -35 region. The greater effect of -10 mutations in general seems to be reflected in the overall distribution of homology coefficients for promoters in Fig. 1(a), where the -35 homology coefficients span twice the range of sigmas that the -10 coefficients do. Increasing a -35 homology from 1 sigma to 5 sigmas may have the same overall effect on promoter activity as increasing a -10 homology from 2 sigmas to 4 sigmas. The parallelogram labelled '[18]' in Fig. 2(b) shows that these 18-spaced promoters behave differently. The -35 region is correlated almost directly with $\log K_B$, whereas an up-mutation in the -10 region should greatly increase $\log k_t$ but slightly decrease $\log K_B$. An example of one such mutation is *lacZp1* going to *lacUV5* (McClure, 1982). The two parallelograms labelled '[17]' and '[18]' in Fig. 2(a) show how the inverse mapping functions T^{-1} , for 17- and 18-spaced promoters, project the black unit square from the isomerization and binding co-ordinates in Fig. 2(b) back on to the -10 and -35 homology system in Fig. 2(a).

It has been shown that 17-spaced promoters are most common (Hawley & McClure, 1983) and are usually stronger than corresponding mutant promoters with different spacings (Ackerson & Gralla, 1982; Stefano & Gralla, 1982; Mulligan *et al.*, 1985). The striking difference in the mathematical behaviour of 17- and 18-spaced promoters is reflected in Fig. 2(b). The 17-spaced promoters from the unit square in Fig. 2(a) are mapped into the upper-right of Fig. 2(b), and the 18-spaced promoters from the same unit square are mapped into the upper-left, giving a horizontal displacement that suggests that the primary effect of spacing is upon the binding constant K_B . This hypothesis is supported by the *tac* promoters in Fig. 1(b), where *tac18* is displaced by -0.7 in $\log K_B$ and $+0.3$ in $\log k_t$ relative to *tac17* (Mulligan *et al.*, 1985). Conversely, the 17-spaced

Fig. 2. Graphic interpretation of the mapping functions T^{-1} and T

(a) The inverse mapping function T^{-1} , which estimates consensus homology from kinetic parameters. The vertical axis is -10 region homology (in sigma units), and the horizontal axis is -35 region homology. The lightly shaded parallelograms in (a) labelled '[17]' and '[18]' show how the T^{-1} matrices for 17- and 18-spaced promoters respectively (eqns. 5 and 8) transform the black unit square in (b) from its own (bind, isom) co-ordinate system into the $(-35, -10)$ system in (a). The [17] parallelogram in (a) shows that 17-spaced promoters from the central region of Fig. 1(b) (with kinetic constants from 7.5 to 8.5 for $\log K_B$ and from -1.5 to -2.5 for $\log k_t$) occupy a broad region towards the left in Fig. 1(a) (with homologies from 1.5 to 3.5 for the -35 region and from 2.5 to 4.0 for the -10 region). The [18] parallelogram in (a) shows that 18-spaced promoters from the same central region of Fig. 1(b) occupy a narrow region towards the right in Fig. 1(a) (with homologies from 3.5 to 4.5 for the -35 region and from 2.5 to 3.5 for the -10 region). (b) The mapping function T , which estimates binding and isomerization kinetics from consensus sequence homology. The vertical axis is the rate of isomerization ($\log k_t$), and the horizontal axis is the strength of binding ($\log K_B$). The lightly shaded parallelograms in (b) labelled '[17]' and '[18]' show how the T matrices for 17- and 18-spaced promoters respectively (eqns. 3 and 6) transform the black unit square in (a) from its own $(-35, -10)$ co-ordinate system into the (bind, isom) system in (b). The [17] parallelogram in (b) shows that 17-spaced promoters from the central region of Fig. 1(a) (with homologies from 3.0 to 4.0 for the -35 region and from 3.0 to 4.0 for the -10 region) occupy a narrow region towards the right in Fig. 1(b) (with kinetic

promoters from the unit square in Fig. 2(b) are mapped on to a broad area to the left in Fig. 2(a), and the 18-spaced promoters occupy a smaller area to the right. The difference in size and shape between the two parallelograms in each Figure suggests that the spacing can also influence how strongly a mutation will affect each of the kinetic constants. It is not clear whether this reflects a fundamental difference due to spacing in the microscopic details of promoter and polymerase conformation, or whether these geometrical shapes eventually will converge as more promoters are analysed kinetically.

Prediction of kinetic constants from sequence homology

To visualize how well the derived equations fit the sample data points (enclosed in ovals or rectangles in Figs. 1a and 1b), these equations were also used to predict the kinetic constants of those promoters from their -10 and -35 homology coefficients in Fig. 1(a). The X symbols (each connected by a line to its promoter) mark the locations in Fig. 1(b) where those promoters are predicted to be according to the mapping equations. Most deviations are within the bounds of the experimental error in measuring the kinetic constants. The T7D, T7C and *Tn10*pIN HH104 promoters do not fit the pattern, and I have no explanation for this discrepancy. One limitation of this analysis is that the 17-spaced promoters are mostly λ and *Tn10* derivatives, whereas the 18-spaced promoters are mostly from *lac* and *gal*, so the observed patterns may be biased somewhat by the limited diversity of the currently available kinetic data set. Insufficient kinetic data exist at the present time to determine the effects of spacings other than 17 or 18 on the mapping function, and it is not clear what kinetic effect there would be for promoters with two possible alternative spacings of almost equal homology, although the *galEpl* promoter (with spacings of either 14 or 21 bases) has a very weak $\log K_B$ of about 6.0 and a strong $\log k_t$ of about -1.5 (Herbert *et al.*, 1986), again suggesting a major effect of spacing on K_B .

Note that the C_B value of 5.5 for the $\log K_B$ of a hypothetical 0.0 sigma 17-spaced 'promoter' and the corresponding C_B value of 4.5 for the 18-spaced promoter agree well with the measured kinetic binding constant of about 4.5 for $\log K_B$ in 'random DNA' (McClure, 1985), since 82% of all possible promoter alignment positions on DNA ('non-specific binding' sites) have homology coefficients that fall between -3 and $+1$ sigmas (Bevington, 1969, pp. 304–309; Studnicka, 1986). The C_t values from these equations also suggest that 'non-specific binding' closed complexes should have a $\log k_t$ of about -5.8 (an isomerization time of about 7 days).

Genetic regulation patterns

Since initiation of transcription involves the two major kinetic steps of binding and isomerization, and since the -35 homology, the -10 homology and the spacing have different effects upon these two steps, it is likely that bacterial cells might have evolved variations in promoter sequences to achieve specific control over K_B and k_t . In

constants from 8.5 to 9.0 for $\log K_B$ and from -1.5 to -2.0 for $\log k_t$. The [18] parallelogram in (b) shows that 18-spaced promoters from the same central region of Fig. 1(a) occupy a broad region towards the left in Fig. 1(b) (with kinetic constants from 7.0 to 8.0 for $\log K_B$ and from -1.0 to -2.0 for $\log k_t$).

Fig. 3. Patterns of differential promoter region homology among different classes of genes and operons

Twelve groups of *E. coli* and bacteriophage promoters are graphed in a format identical with that of Fig. 1(a). The vertical axis in each rectangular graph represents consensus homology (in sigma units) for the -10 promoter region, and the horizontal axis is for the -35 region. Each black dot represents a particular promoter from Fig. 1(a). The non-random clustering of a class of promoters in each small graph above (and thus in Fig. 1a) suggests that *E. coli* and its bacteriophages might have evolved differential strength in either the -10 region or the -35 region to achieve greater control over the slow step of transcription initiation. (a) All types of promoters from Fig. 1(a); (b) promoters governing operons for catabolic pathways regulated by CRP (cyclic-AMP-receptor protein) (deCrombrughe *et al.*, 1984); (c) promoters for regulatory proteins such as repressors and activators; (d) promoters for ribosomal protein operons; (e) promoters for stable RNA operons, such as tRNA and rRNA; (f) stringently controlled promoters, responding to amino acid starvation (Sarmientos & Cashel, 1983; Travers, 1984); (g) promoters for amino acid biosynthesis operons; (h) promoters from fd bacteriophage, a filamentous single-stranded DNA phage related to M13 phage, which exhibits no temporal phases of gene expression and which does not kill the host cell; (i) promoters from ϕ X, G4 and T7 bacteriophages, which are virulent phages that kill the host cell; (j) promoters for *E. coli* genes that respond to heat shock; (k) promoters from T5 bacteriophage, another virulent phage; (l) promoters from λ bacteriophage, a temperate phage with alternate cycles of virulence and lysogeny (white dots denote mutant λ promoters).

Fig. 3, 12 classes of promoters from Fig. 1(a) are shown separately, in greatly reduced format, so that the patterns of non-random clustering are easily visualized. Cyclic-AMP-receptor protein-controlled promoters (Fig. 3b) and promoters for repressors and activators (Fig. 3c) have weak -35 regions and varied -10 regions. Promoters for ribosomal proteins (Fig. 3d), for stable

RNA (Fig. 3e) and for stringent-response genes (Fig. 3f) have strong -10 regions with varied -35 regions. Promoters for amino acid-biosynthetic operons (Fig. 3g) have weak -10 regions and varied -35 regions. Filamentous fd bacteriophage (Fig. 3h) promoters have strong -10 regions with weak -35 regions, whereas ϕ X, G4 and T7 bacteriophage (Fig. 3i) promoters are the opposite, with weak -10 regions and strong -35 regions. T5 bacteriophage (Fig. 3k) has two distinct types of promoters, some with very strong -10 regions and moderate -35 regions, and others with very strong -35 regions and moderate -10 regions. Heat-shock-gene promoters (Fig. 3j) have extremely weak -10 regions with moderate -35 regions, reflecting the role of the RpoH protein in recognizing their altered -10 regions. Temperate λ bacteriophage (Fig. 3l) has a variety of different kinds of promoters that do not strongly cluster on the graph. Thus two different kinds of kinetic control might be achieved by varying the degree of homology in either the -10 or the -35 region. Of course, it is too simplistic to assume that variations in binding and isomerization kinetics could account for all of the regulatory patterns of every gene. However, the analysis in Fig. 3 suggests that homology-dependent kinetic effects might contribute in a consistent way to a more complex overall regulatory scheme.

Conclusion

Two linear equations have been derived that map (H_{-35}, H_{-10}) homology coefficients on to $(\log K_B, \log k_t)$ kinetic coefficients, giving support to a modified version of the bipartite model. Also, specific classes of genes seem to utilize differential strength in the -35 and the -10 regions to achieve control over the rate-limiting step of transcription initiation. However, the small number of promoters that so far have been subjected to the abortive initiation assay places a limit upon the interpretability of these results, and more refined estimates of the coefficients in the mapping function must await continued kinetic experiments. As the kinetic effects of more individual nucleotide substitutions in the promoter site are examined, the homology matrices (obtained from nucleotide frequencies at each position of the promoter consensus) could eventually be replaced by matrices reflecting the kinetic parameters K_B and k_t associated with each mutation, and these new matrices then could be used to predict reliable values for K_B and k_t directly from the DNA sequences of previously uncharacterized putative promoter sites. However, the kinetic activities of many mutant promoters remain to be investigated, and the fine structure of each one is unique, with weaker homologies often overlapping the dominant one (Malan & McClure, 1984; Mulligan *et al.*, 1985; Herbert *et al.*, 1986). Making identical base substitutions at the same position on two different promoters might not have consistent effects on the kinetic constants, since one of the weaker overlapping homologies might be intensified to yield heterogeneous open complexes (Mulligan *et al.* 1985), which might interfere with binding or isomerization. For example, *lac*L157 is a single-point mutation that falls in the -35 region of the *lacZp1* promoter and also in the -10 region of the overlapping *lacZp2* promoter (Malan & McClure, 1984), with complex kinetic effects. It is easy to predict successfully that very strong homologies will be functional. However, some biologically active promoter homologies are very

weak (Studnicka, 1987). What biochemical mechanisms determine that certain weak homologies are able to function and that others are not?

REFERENCES

- Ackerson, J. W. & Gralla, J. D. (1982) Cold Spring Harbor Symp. Quant. Biol. **47**, 473–476
- Bachmann, B. J. (1983) Bacteriol. Rev. **47**, 180–230
- Bevington, P. R. (1969) Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York
- Cowing, D. W., Bardwell, J. C. A., Craig, E. A., Woolford, C., Hendrix, R. W. & Gross, C. A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 2679–2683
- Dayton, C. J., Prosen, D. E., Parker, K. L. & Cech, C. L. (1984) J. Biol. Chem. **259**, 1616–1621
- deCrombrughe, B., Busby, S. & Buc, H. (1984) Science **224**, 831–838
- Fournier, M. J. & Ozeki, H. (1985) Microbiol. Rev. **49**, 379–397
- Gentz, R. & Bujard, H. (1985) J. Bacteriol. **164**, 70–77
- Gilbert, W. (1976) in RNA Polymerase (Losic, R. & Chamberlin, M., eds.), pp. 193–205, Cold Spring Harbor Press, Cold Spring Harbor
- Hawley, D. K. & McClure, W. R. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 6381–6385
- Hawley, D. K. & McClure, W. R. (1983) Nucleic Acids Res. **11**, 2237–2255
- Herbert, M., Kolb, A. & Buc, H. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 2807–2811
- Johnston, D. E. & McClure, W. R. (1976) in RNA Polymerase (Losic, R. & Chamberlin, M., eds.), pp. 413–428, Cold Spring Harbor Press, Cold Spring Harbor
- Lindahl, L. & Zengel, J. M. (1986) Annu. Rev. Genet. **20**, 297–326
- Malan, T. P. & McClure, W. R. (1984) Cell **39**, 173–180
- McClure, W. R. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 5634–5638
- McClure, W. R. (1982) Cold Spring Harbor Symp. Quant. Biol. **47**, 477–481
- McClure, W. R. (1985) Annu. Rev. Biochem. **54**, 171–204
- 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
- Nomenclature Committee of the International Union of Biochemistry (NC-IUB) (1985) Biochem. J. **229**, 281–286
- Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 784–788
- Sarmientos, P. & Cashel, M. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 7010–7014
- Simons, R. W., Hoopes, B. C., McClure, W. R. & Kleckner, N. (1983) Cell **34**, 673–682
- Stefano, J. E. & Gralla, J. D. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 1069–1072
- Studnicka, G. M. (1986) Comput. Appl. Biosci. **2**, 269–275
- Studnicka, G. M. (1987) Gene **58**, 45–57
- Travers, A. A. (1984) Nucleic Acids Res. **12**, 2605–2618
- Yu, X. M. & Reznikoff, W. S. (1986) J. Mol. Biol. **188**, 545–553

Received 1 September 1987/10 November 1987; accepted 1 March 1988