Helix Stability in Prokaryotic Promoter Regions

Hanah Margalit,*,‡ Bruce A. Shapiro,§ Ruth Nussinov,§,∥ John Owens,§ and Robert L. Jernigan‡

Laboratory of Mathematical Biology, National Cancer Institute, National Institutes of Health, Building 10, Room 4B-56, Bethesda, Maryland 20892, Laboratory of Mathematical Biology, National Cancer Institute, Frederick Cancer Research Facility, Building 469, Room 151, Frederick, Maryland 21701, and Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv, Israel 69978

Received October 19, 1987; Revised Manuscript Received February 19, 1988

ABSTRACT: Prokaryotic promoters have been extensively studied to relate sequence features to promoter function. Here we examine the relationship between double-helix stability and promoter activity. The double-helix stability is evaluated from sequence data by free energy computation, based on reported values of dinucleotide free energies for strand separation. For a collection of 168 promoters, we find that within a 500-nucleotide span around the transcription initiation site the −10 region is the least stable. There is no correlation between the free energies and the rates of RNA polymerase–promoter open complex formation measured for 25 promoters. We also compare the free energies of 121 promoter mutations across the −35 and −10 consensus regions with the free energies of the corresponding wild-type sequences. These pairwise mutant–wild-type comparisons provide a particularly good test since the examined sequences differ only in one nucleotide so that all other sequence-dependent effects remain the same. About 80% of the mutations in the −10 region that show increased/reduced promoter activity are less/more stable than the wild types. The observed high free energy peak and the mutation data strongly support the conjecture that the instability, or melting properties, of the −10 region plays a significant role in promoter function.

Sequence comparisons of *Escherichia coli* promoter regions have shown two conserved hexamers: TATAAT, located about 10 base pairs upstream from the transcription start position (the −10 consensus sequence), and TTGACA, located about 35 base pairs upstream from the transcription start position (the −35 consensus sequence). The importance of these two consensus sequences has been established by genetic and biochemical means. Most mutations that affect promoter activity fall in these regions. Mutations that reduce the homology with the consensus decrease promoter activity (down mutations), while mutations increasing the homology with the consensus show increased promoter activity (up mutations) [reviewed by von Hippel et al. (1984) and McClure (1985) and compiled by Hawley and McClure (1983) and Harley and Reynolds (1987)]. In addition, "dimethyl sulfate probe" experiments show that most of the contacts between the promoter and the RNA polymerase occur in or near these regions (Siebenlist et al., 1980).

The consensus sequences are composed of nucleotides that are the most frequent at specific positions in the promoter region. However, no wild-type promoter has been discovered with the two exact consensus sequences, and there seems to be a large variability among individual promoter sequences. Despite this variability, all are functional as the sites at which RNA polymerase interacts specifically to initiate transcription, and thus they must share common features that enable them to serve as suitable sites for transcription initiation. Indeed, there are studies showing a common pattern in the helical angles in the promoter region (Nussinov, 1984), suggesting that conformational features are important for the specific interaction with RNA polymerase. Other studies show deviations from standard DNA conformations that might serve as a signal for RNA polymerase binding (Nussinov, 1985). In the present study, such issues will not be considered; we are examining the postulate that the local instability of the double helix is a common feature of prokaryotic promoters.

There is supporting evidence that *E. coli* RNA polymerase interacts with the promoter in two steps to initiate transcription. The first step is the formation of an inactive "closed" complex between the enzyme and the double-stranded DNA. The second step is the isomerization of the closed complex into an active "open" complex [reviewed by von Hippel et al. (1984) and McClure (1985)]. The second step, termed "melting in", has been observed to be temperature and salt dependent [reviewed by von Hippel et al. (1984)] and clearly involves melting of the DNA. Also, reagents that decrease the T_m of DNA increase the efficiency of transcription initiation (Nakanishi et al., 1974). This might suggest that ease of melting or strand separation of the DNA in the promoter region is an important feature for promoters. It has been shown that transcription initiation involves unwinding of a short segment of DNA, from the middle of the −10 region to the transcription start position (Wang, 1982). "Chemical probe" experiments show that the minimum unwound region extends from −9 to +3 (Siebenlist et al., 1980). A possible model is for the RNA polymerase to interact with base pairs upstream from the unwound region to form the closed promoter complex. This binding would provide a high concentration of RNA polymerase near the segment that will be unpaired in the open complex and place the RNA polymerase in a position to facilitate its interaction with the −9 to +3 region when unpaired by thermal fluctuations (von Hippel et al., 1984; Gilbert, 1976). Since the −9 to +3 region does not seem to be exceptional in its A + T content, it was not clear whether the "melting in" at that region depended on the instability of the DNA or was only a result of the interaction with the relevant segment of the enzyme.

Our aim here is to evaluate the stabilities of prokaryotic promoter sequences. A recent calorimetric study by Breslauer et al. (1986) provides the means for a numerical evaluation

‡ National Institutes of Health.
§ Frederick Cancer Research Facility.
∥ Tel-Aviv University.

of the stability of double-stranded DNA segments. We are evaluating the free energies or stabilities of short segments within the promoter region, using the dinucleotide transition free energies reported by Breslauer et al. (1986). A somewhat similar study was carried out by Gotoh and Tagashira (1981b), who analyzed the melting profiles of a few prokaryotic sequences and showed that the promoters tend to fall in relatively unstable DNA regions. We show for a large database of 168 prokaryotic promoters that the −10 region has a statistically significant higher free energy (is less stable) than the neighboring regions. This follows directly from the unusually high frequency of the doublet TA in the promoter region, which has the highest free energy value of all dinucleotides. We also calculate the free energy change for 121 up and down mutations spread across the −10 and −35 consensus regions and show that there is a correlation between the direction of free energy change and the type of mutation in the −10 region, but not in the −35 region. In addition, there are reported values for the rate of open complex formation for about two dozen promoters. We use these values to examine whether there is a correlation between the stability of the DNA in the promoter region and the rate of open complex formation.

Materials and Methods

Free Energy Calculation. A general formula (Breslauer et al., 1986) for calculating the transition free energy from double helix to single strand of a DNA sequence n nucleotides long, ΔG_n, is

$$\Delta G_n = -(\Delta g_{\text{ini}} + \Delta g_{\text{sym}}) + \sum_{i=1}^{n-1} \Delta g_{i,i+1}$$

where Δg_{ini} is the initiation free energy (6 kcal/mol for a sequence without any G or C and 5 kcal/mol otherwise). Δg_{sym} accounts for the entropic difference between a DNA double strand formed from a self-complementary sequence and a double strand formed from a sequence that is not self-complementary (RT ln 2 kcal/mol and 0 kcal/mol, respectively). Δg_{ij} is the transition free energy for dinucleotide of type ij. Because it is usual to refer to the lower free energy state as the more stable state, we have found it more convenient to use the negatives of their values, i.e., to refer to the free energy decrease on passing from coil to helix. Thus, the Δg_{ij}'s used are the negatives of those reported by Breslauer et al. (1986) at 25 °C (in kcal/mol):

3′ \ 5′	A	C	G	T
A	−1.9	−1.9	−1.6	−0.9
C	−1.3	−3.1	−3.1	−1.6
G	−1.6	−3.6	−3.1	−1.9
T	−1.5	−1.6	−1.3	−1.9

Results

Relative Stability of the Promoter Region. (A) Database. Our data include 168 *E. coli* promoter sequences and their flanking sequences (not more than 250 nucleotides upstream and downstream from the transcription start position). The data are based on the most recent compilation of *E. coli* promoters (Harley & Reynolds, 1987) and on the information that appears in GENBANK. The sequences have been derived from GENBANK (release 52) and are listed in the supplementary material (see paragraph at end of paper regarding supplementary material). We include in our database only sequences for which the transcription start positions have been identified experimentally. Most of the promoters in the database interact with *E. coli* σ^{70}, the factor that is essential for the recognition of the majority of promoters interacting with *E. coli* RNA polymerase. Since we are not studying here interactions between specific residues of the holoenzyme and specific nucleotides in the promoter sequence, we do not divide the database into subgroups according to the species of the RNA polymerase holoenzyme interacting with them. The local instability of the DNA is a property that might appear in all promoters. The analysis that we perform is statistical and as such can detect a general tendency of the promoters to be locally less stable. If a minority of the promoters are exceptional and belong to a specific biologically defined group, this analysis would not find it. Also, if this feature exists only within a specific group of promoters, the statistical analysis performed on the whole database would not detect it.

(B) Statistical Analysis. Our purpose is to examine the distribution of free energies along the sequences and to test the hypothesis that the free energy in the promoter region is higher than in the rest of the sequence. Although individual sequences might show different and conflicting patterns, we are interested in showing that the low stability is statistically significant in the "promoter population" since our sample should be large enough to detect such characteristics if they exist.

Each sequence is divided into overlapping blocks of length l. The kth block extends from base pair k through base pair $k + l - 1$. We calculate the free energy for each block as explained above. Because we deal with short blocks within a long sequence, we omit the Δg_{sym} term. Also, no significant difference has been observed between the free energy distributions obtained with and without taking Δg_{sym} into account (because the fraction of symmetric blocks in our database is very small; 0.1% when $l = 12$ and 2% when $l = 6$). All the sequences are aligned at the transcription start position. No gaps are introduced in order to align the −10 consensus regions or the −35 consensus regions. We add the free energies of the blocks at each position over all the sequences and normalize it by the number of entries at each position. The end result of this process is a distribution of free energies of blocks l nucleotides long around the transcription start position that includes contributions of all the sequences in the database. It is possible either to examine the distribution of the block free energies or to calculate the mean of the distribution and to examine the blocks according to their deviations from the mean. Breslauer et al. (1986) have shown that there is a remarkably good fit of their calculations to experimentally measured free energies of oligonucleotides of length 6–14; we do not use block lengths beyond that range. Because it has been shown experimentally that the minimum unwound region is 12 nucleotides long (Siebenlist et al., 1980), we have initially chosen l to be 12. Figure 1 shows a plot of the deviations from the mean of blocks of length 12. There is a peak in free energies in the −10 region but not in the −35 region.

To test the significance of the results, we compare the blocks at each relevant position to a randomly chosen sample of blocks of the same length. The random sample is derived by applying a function that chooses randomly one block of length l in each sequence. Each sequence is presented once in the promoters' group by its free energy value for the relevant block and once in the random group by a free energy value of a randomly chosen block. We perform a median test, that provides the same type of evaluation as a t test but does not assume a normal distribution, to test the hypothesis that the group of blocks starting at a relevant position within the promoter region has a higher free energy value than the random sample. For $l = 12$, the most significant results are achieved for the blocks starting at −19, −18, ..., −10 and extending 11 nucleotides downstream ($p < 0.001$). No significant deviation from

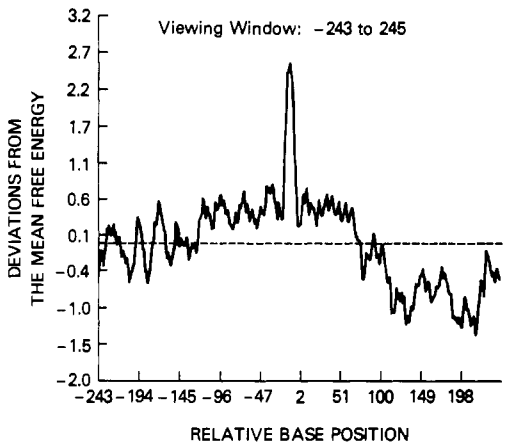

FIGURE 1: Distribution of free energies around the transcription start position. The graph shows the result of summing and normalizing the dinucleotide free energy values for 168 *E. coli* promoter regions. The graph is centered at the transcription start position (position 0) and extends for 250 nucleotides on either side of position 0. Breslauer's free energy values are used with sliding blocks of 12 bases. The dashed line represents the mean value of the normalized free energy sum. The abscissa is the nucleotide position relative to the transcription start position. The ordinate is the deviation from the mean free energy (kilocalories per mole). Note the peak within the -10 consensus region.

randomness is found in the -35 region.

Is the observed peak in free energy around the -10 region a reflection only of the composition, the high proportion of A's and T's, or does it result from the neighbor dependence of the base pairs in the sequence? From the table of dinucleotide transition free energies, it is clear that it is possible to have sequences composed of just A's and T's with lower free energies than some sequences that include special combinations of A, C, G, and T. For example, the sequence AAAAAAAAAAAA which is entirely A has a free energy value of -14.9 kcal/mol while the free energy of a sequence which is only 50% A like AGAGAGAGAGAG is -12.6 kcal/mol. Another example provides the sequences AATAATAATAAT (-10.3 kcal/mol) and ACTACTACTACT (-9.3 kcal/mol). Figure 2a–d shows the frequency distribution of AA, AT, TA, and TT around the 168 transcription start positions. Each position represents a midpoint of a block of 12 and shows the frequency of the specific dinucleotide in the corresponding block. There is no peak in AA frequency in the promoter region (Figure 2a) or in TT frequency (Figure 2d) except a peak around the -35 region which results from the -35 consensus sequence *TT*GACA. There is a strong peak in the TA frequency (Figure 2c) and a somewhat weaker one in AT frequency (Figure 2b) around the -10 region (as a result of the -10 consensus sequence *TA TA AT*). Recall that TA has the highest free energy (-0.9 kcal/mol) and the free energy of AT is also fairly high (-1.5 kcal/mol), and hence, these two dinucleotides are major contributors to the peak in the free energy shown in Figure 1. We have examined the frequency distribution of the other 12 possible dinucleotides. We find that TC and GT also appear to be abundant in the -10 region. These dinucleotides also have relatively high free energies (-1.6 and -1.3 kcal/mol, respectively) and also contribute to the free energy peak in Figure 1.

The region that has been reported to be unwound is -9 to +3 (Siebenlist et al., 1980). The free energy of the blocks of length 12 starting at -9 is still significantly higher than the random sample values ($p < 0.01$), but not as highly significant as for the other blocks in the -10 region. The blocks starting at -8 do not show significant deviation from randomness. Repeating the analysis with block lengths of 6 shows that the

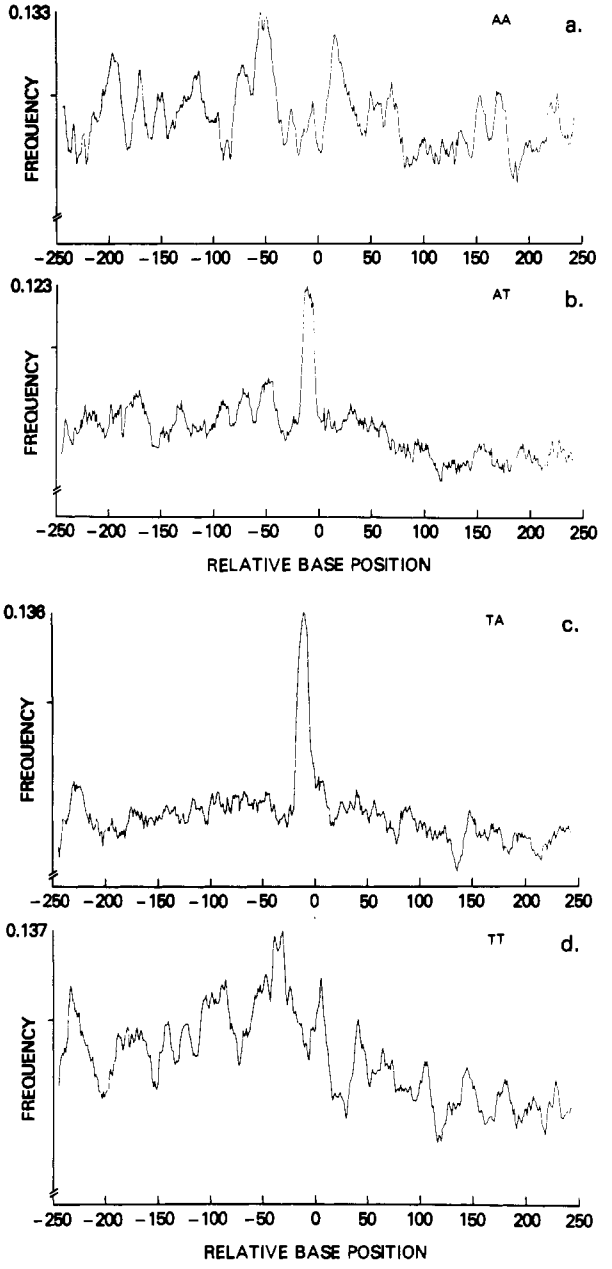

FIGURE 2: Distribution of dinucleotides around the transcription start position. The sequences in the database are aligned at the transcriptional start position (designated as position 0). For each position, we calculate the frequency of each of the 16 possible dinucleotides. The plots illustrate the dinucleotide frequencies in overlapping blocks of length 12. The frequencies are normalized by the number of entries in each block. Every point represents a midpoint of a block of 12 nucleotides. The frequency distributions of the four dinucleotides most relevant to our study are shown: (a) AA; (b) AT; (c) TA; (d) TT.

blocks starting at positions -9, -8, and -7 have statistically significant higher free energies in comparison to random samples ($p < 0.001$ to $p < 0.05$). The blocks starting at -6, -5, -4, and -3 do not have higher free energy values in comparison to the rest of the sequence. The analysis for $l = 6$ is illustrated in Figure 3.

(*C*) *Melting Temperature Comparisons.* From the values of dinucleotide enthalpies and entropies reported by Breslauer et al. (1986), it is possible to derive the melting temperatures of the dinucleotides. This set of values does not correlate well with two other sets of dinucleotide melting temperatures reported previously by Gotoh and Tagashira (1981a,b). The differences might be due to the variations in the salt concentrations at which the measurements were performed [0.019

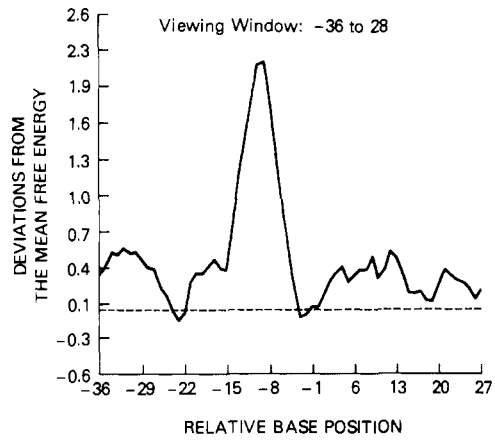

FIGURE 3: Distribution of free energies around the transcription start position. This graph utilizes the same technique for its generation as described in Figure 1 except that a sliding block of size 6 is used and a smaller portion of the original viewing window is depicted to allow closer scrutiny. A major peak occurs around the −9 position, concurring with expectations based on experimental evidence of unwinding.

M (Gotoh & Tagashira, 1981a), 0.19 M (Gotoh & Tagashira, 1981b), and 1 M (Breslauer et al., 1986)]. The analysis represented by Figure 1 has been repeated using the values of the melting temperatures (not shown). The block value now is the average of the dinucleotide melting temperatures rather than the free energy value. This average is not intended to be an estimate of the actual melting temperature of the block but merely a way to compare results for the three different sets of dinucleotide melting temperatures. Although the three sets are different, the three plots using the melting temperature data show a minimum in the −10 region and have a similar appearance. We think that this occurs because of the high frequency of TA in this region and because TA has the lowest melting temperature for all three sets.

Stability of Mutated Promoter Sequences. Are up/down promoter mutations less/more stable (with higher/lower free energy values) than the wild-type sequences? To answer this question, we compare the free energies of promoter sequences including a single point mutation with the corresponding wild-type sequences.

(*A*) *Database.* Our database of reported prokaryotic promoter mutations includes 50 down mutations and 18 up mutations in the −10 region, and 44 down mutations and 9 up mutations in the −35 region. Of the total of 121 mutations, 96 are located exactly in the −10 and −35 hexamers as they are aligned according to the published compilations (Hawley & McClure, 1983; Harley & Reynolds, 1987). The other mutations are located extremely close to the consensus hexamers, with 1 mutation located 1 nucleotide downstream from the −10 hexamer, 21 mutations located within a distance of 3 nucleotides, and 3 mutations located 4 nucleotides upstream from the consensus hexamers. Only single point mutations are included. We do not distinguish among organisms but include all in one group.

(*B*) *Free Energy Comparisons.* In the nearest-neighbor approximation, a point mutation at position k could change the free energy values of the dinucleotides $(k-1, k)$ and $(k, k+1)$. The values of other dinucleotides in the relevant region remain constant. Therefore, we simply calculate the sum of free energies of the two dinucleotides that include the mutation. We assume that Δg_{ini} is constant and do not include it. Theoretically, it is possible that a mutation can change a region without any G or C ($\Delta g_{ini} = -6$ kcal/mol) to one that includes a G or a C ($\Delta g_{ini} = -5$ kcal/mol) and vice versa, but this is an unlikely event. For example, if we consider segments of length 12 around the mutation, the probability of having such a segment with no more than 1 G or C is very low. In order to confirm that treating Δg_{ini} as a constant is a resonable assumption, we have repeated the calculations examining a block of 12 rather than a triplet just for the −10 region mutations. There is no significant change in the results with and without Δg_{ini} included.

The free energy differences between the mutated promoters and the corresponding wild-type sequences are listed in Table I and summarized in Table II. The significance of the results is evaluated by a Wilcoxon matched pairs test.

In 40 out of 50 down mutations in the −10 region, the free energies have decreased in comparison to the wild types. This result is statistically highly significant ($p < 0.00003$). Furthermore, it is notable that 21 of these 40 mutations have a free energy change greater than 1 kcal/mol which is a strong effect since thermal energy is only about 0.6 kcal/mol. Note that 8 of the 40 mutations are transversions of A to T and T to A; without taking the nearest-neighbor effect into account, those would have been expected to have no effect on the DNA stability. For seven of these eight transversions, the free energy differences are greater than 1 kcal/mol. Out of the 18 up mutations, 14 show increased free energies in comparison to the wild-type sequences ($p < 0.005$). For 8 of these 14, the free energy increase is more than 1 kcal/mol.

Only 21 out of the 44 down mutations in the −35 region appear to be more stable than the wild types, and only 5 out of the 9 up mutations in that region are less stable. These results are not statistically significant.

Correlation between Helix Stability and Rate of Open Complex Formation. It has been shown that the rate of isomerization from a closed complex of promoter–RNA polymerase to an open active complex can be evaluated by using the abortive initiation reaction (McClure, 1980). The rates of open complex formation have been reported for about 25 promoters (McClure, 1980; Dayton et al., 1984; Hawley & McClure, 1980, 1982; Stefano & Gralla, 1982; Malan et al., 1984; Shih & Gussin, 1983; Simons et al., 1983; Mulligan et al., 1985) and can be used for examining the correlation between the stability of the promoter region and the rate of open complex formation. Because the kinetic measurements of the rate constants were performed at 37 °C, we have used for this analysis free energy values corrected to 37 °C according to the data reported by Breslauer et al. (1986). We have calculated the correlation coefficient between the rate of open complex formation and $e^{-\Delta G/RT}$ for segments of different lengths within the promoter region. No significant correlation has been found either in the −10 region or in the −35 region. We will further discuss this result below.

Discussion

Prokaryotic promoter regions have been considered as relatively unstable, especially in the −10 consensus region, because of the high frequency of A's and T's within them. A recent study by Breslauer et al. (1986) enables us to better predict the stability of a DNA double-helical segment by using neighbor-dependent free energy values. These values, however, are unlikely to include all sequence-dependent effects. Those which appear to be longer range in character are clearly not taken into account. By using their data, we are able to quantitatively compare the stability of segments within the promoter region with other DNA segments by simple free energy calculations. In a statistical analysis of 168 prokaryotic promoters, we show that the −10 region is locally unstable. This result is statistically highly significant ($p < 0.001$). We

have compared the free energy values of blocks of length 12 within each promoter with the free energies of all blocks of the same length in its flanking regions (not shown). In about 60% of the sequences, the blocks representing the −10 region are in the high 30% of block free energy values, and about 30% of the promoters rank in the high 10%. The striking free energy peak in the −10 region (Figure 1) results from the specific order of A's and T's in the −10 consensus sequences. We find that the helically least stable TA dinucleotide is very frequent in the −10 region in comparison to other DNA segments. Assuming independence between neighboring nucleotides, we can calculate the expected frequency for a specific dinucleotide at each position by multiplying the frequencies of the individual nucleotides. Comparison of the observed and expected frequencies averaged over blocks of 12 nucleotides indicates that the observed frequency of TA in the flanking regions is lower and its frequency in the −10 region is higher than expected. Thus, one of the roles played by TA in this region might be to facilitate the "melting in" step.

The free energy value of the 12-nucleotide block starting at −9 is significantly higher than the free energies of similar-size blocks in a random sample, but not as highly significant as for the upstream blocks (−19, −18, ..., −10). Repeating the analysis with a block length of 6 shows that the blocks starting at positions −9, −8, and −7 are relatively unstable. The stabilities of other blocks within the −9 to +3 segment do not appear to be different from the rest of the sequence (Figure 3). It is possible that the "opening" at the 5′ end of the −9 to +3 segment is followed by a "zipper" effect corresponding the the cooperative opening of a longer segment.

Studies on prokaryotic promoter mutations have revealed that most of the down mutations are those with decreased homology to the consensus promoter while most of the up mutations show increased homology (von Hippel et al., 1984; McClure, 1985; Hawley & McClure, 1983). We examine 121 up and down mutations in the two consensus regions by comparing their free energy values with those of the corresponding wild-type sequences. If the ease of melting plays a role in the steps preceding transcription initiation, we would expect the down/up mutation sequences to have lower/higher free energy values than the wild-type sequences. Without taking the nearest-neighbor effect into account, just by considering the type of mutation, about 63% of the mutations in the −10 region would have been predicted to have the expected effect on the DNA stability. Our analysis which considers nearest-neighbor effects shows that in the −10 region 80% of the down mutations are more stable than the corresponding wild-type sequences and 77% of the up mutations are less stable. Since T $\rightleftarrows$ A transversions do not change the A + T content, they have been interpreted as affecting the specific contacts with the enzyme rather than changing the melting properties of the DNA. While this might still be the case, we also show that 8 out of 10 such transversions among the down mutations in the −10 region show decreased free energies. There are only three such transversions among the up mutations in the −10 region. For two out of the three, there is no free energy change, and one shows an increased free energy value. In the −35 region, there is no significant correlation between the direction of free energy change and the mutation type. We also do not find the −35 region to be relatively unstable. These findings support the model suggested by Gilbert (1976), namely, that the −35 region is recognized by the RNA polymerase and the −10 region is responsible for the "melting in" step.

The analyses performed here are statistical in nature. Clearly, a correspondence between two properties, even statistically significant, does not necessarily imply functional correlation but might be due to other factors. In our case, the −10 consensus sequence could be essential for specific contacts with the enzyme, and, accidentally, because of its richness in the relatively unstable dinucleotides TA and AT, the helix instability appears to be correlated with promoter activity. In order to control this indirect effect as much as possible, we have done the following: We retain all the −10 region wild-type triplets for which a mutation in the middle nucleotide could either increase or decrease the free energy. We exclude those triplets for which all possible mutations would yield the same directional change in the free energy. This smaller database includes 19 sequences (11 down and 8 up mutations, indicated in Table I by an asterisk); 7 out of the 11 down mutations show decreased free energy, and 8 out of the 8 up mutations show increased free energy. Hence, out of a total of 19 mutations, 15 show a free energy change that would have been expected if there was a correlation between the helix stability and transcription activity. If the mutations were random, we would expect only 10 out of the 19 mutations (53%) to show a free energy change in this direction. The expected frequency under randomness is calculated simply by taking into account for each triplet in this smaller database all the possible substitutions in the middle nucleotide. The difference between the expected (assuming random mutations) and observed is statistically significant ($p < 0.05$ by a χ^2 test). The changes in free energy deviate significantly from randomness ($p < 0.005$ by a Wilcoxon test).

Another way to control for this possible indirect effect is the following: exlcusion of all the mutations in positions with known interactions with the enzyme and retaining a database of mutations in positions that were not identified to be in specific contacts with the enzyme. According to Siebenlist et al. (1980), the 5′-T and the 3′-T of the −10 consensus sequence (*TATAAT*) were not found to interact specifically with the enzyme. These two nucleotides are highly conserved and probably have a role in promoter function. There are 18 down mutations in these positions, and 14 of those show decreased free energy; 5 up mutations are located in these positions, and all of them show increased free energy. The statistical significance of these results is $p < 0.005$ by a Wilcoxon test. The importance of the T's in these positions might be in their influence on the helix stability. These two more limited analyses also reinforce the conclusion regarding the role of the helix stability in promoter activity.

The mutation data enable a comparison of free energies between a mutant and its corresponding wild type. This pairwise comparison suggests a direct relationship between transcription activity and the degree of helix instability at the −10 region. As usually formulated, the two steps leading to open complex between the enzyme and the promoter are

$$P + R \overset{K_B}{\rightleftharpoons} RP_c \underset{k_r}{\overset{k_f}{\rightleftharpoons}} RP_o$$

P is the promoter, R is the RNA polymerase, K_B is the equilibrium constant of the closed complex formation, k_f is the rate of isomerization from a closed (RP_c) to an open (RP_o) complex, and k_r is the rate of the reverse process. Since the second step involves DNA melting, it is conveivable that the degree of instability could affect this step.

An explanation for such a direct relationship between the helix stability and the process of open complex formation can be found if the above scheme is expanded in more detail in the following way (von Hippel et al., 1984): (1) The RNA polymerase interacts with the −35 region to form the closed

Table I: Free Energy Differences between Promoter Mutations and the Corresponding Wild-Type Sequences[a]

WT sequence	mutant sequence	free energy difference (kcal/mol)	description of mutation: gene	position	ref
			−10 Region Down Mutations		
*AAG	ACG	−1.4	λ PRE	−11	Wulff et al. (1980)
*AAG	AGG	−1.2	λ PRE	−11	Wulff et al. (1984)
*AAG	ATG	0.1	λ PRE	−11	Wulff et al. (1984)
ACT	ATT	−0.5	gal P2[b]	−8	Bingham et al. (1986)
AGT	AAT	−0.5	arg	−12	Piette et al. (1982)
			trp	−14	Oppenheim et al. (1980)
AGT	ATT	−0.5	Tn10 tetA	−13	Daniels & Bertrand (1985)
ATA	AAA	−1.4	P22 ant	−10	Youderian et al. (1982)
ATA	ACA	−0.8	tyrT	−13	Berman & Landy (1979)
			P22 ant	−10	Youderian et al. (1982)
			λ cin	−12	Rosenberg et al. (1978)
ATA	AGA	−0.8	P22 ant	−10	Youderian et al. (1982)
ATG	AAG	−0.1	tyrT	−8	Berman & Landy (1979)
ATG	ACG	−1.5	gal P1	−15	Busby et al. (1982)
			tyrT	−8	Berman & Landy (1979)
			lac P1	−10	LeClerck & Istock (1982)
ATT	AAT	0	λ C17	−7	Mozola et al. (1979)
ATT	ACT	0.5	str	−6	Post et al. (1978)
			λ PRE	−7	Wulff et al. (1980)
CTA	CAA	−1.3	P22 ant	−12	Youderian et al. (1982)
CTA	CCA	−2.5	P22 ant	−12	Youderian et al. (1982)
			gal P2	−7	Bingham et al. (1986)
GTA	GCA	−2.8	lac P1	−12	LeClerck & Istock (1982)
GTA	GGA	−2.5	Tn5 NPTII	−11	Rothstein & Reznikoff (1981)
*GTT	GCT	−1.5	trp	−13	Oppenheim et al. (1980)
					Miozzari & Yanofsky (1978)
TAA	TGA	−0.7	Tn10 tetR	−11	Daniels & Bertrand (1985)
TAC	TGC	−2.8	leu	−12	Gemmill et al. (1984)
			λ cin	−11	Rosenberg et al. (1978)
TAC	TTC	−1.3	ara B	−13	Wilcox et al. (1982)
TAT	TCT	−0.8	lac P1	−11	LeClerck & Istock (1982)
TAT	TGT	−0.8	λ C17	−11	Mozola et al. (1979)
			P22 ant	−11	Youderian et al. (1982)
			P22 ant	−9	Youderian et al. (1982)
			lac P1	−11	LeClerck & Istock (1982)
					Reznikoff et al. (1982)
TAT	TTT	−1.4	P22 ant	−9	Youderian et al. (1982)
			P22 ant	−11	Youderian et al. (1982)
			lac P1	−11	LeClerck et al. (1984)
			gal P1	−11	Musso et al. (1977)
TGA	TAA	0.7	λ PRM	−14	Rosen et al. (1980)
			arg	−16	Piette et al. (1982)
TGC	TAC	2.8	gal P1	−14	Busby et al. (1984)
*TGT	TAT	0.8	λ PRE	−14	Wulff et al. (1980)
			gal P2	−14	Bingham et al. (1986)
*TGT	TTT	−0.6	gal P2	−14	Bingham et al. (1986)
TTA	TCA	−0.7	Tn10 tetA	−6	Daniels & Bertrand (1985)
TTA	TGA	−0.7	Tn10 tetA	−11	Daniels & Bertrand (1985)
*TTC	TAC	1.3	P22 ant	−7	Youderian et al. (1982)
*TTC	TCC	−1.2	P22 ant	−7	Youderian et al. (1982)
*TTG	TCG	−1.4	tyrT	−16	Berman & Landy (1979)
			lac P1	−7	LeClerck et al. (1984)
			−10 Region Up Mutations		
*ACA	ATA	0.8	ampC	−11	Jaurin et al. (1982)
AGT	AAT	−0.5	fol	−8	Smith et al. (1982)
ATT	AAT	0	P22 ant	−8	Grana et al. (1985)
*CCA	CTA	2.5	λ cin	−7	Rosenberg et al. (1978)
			trp hyb	−7	Oppenheim & Yanofsky (1980)
*CCC	CTC	3.0	trp hyb	−8	Oppenheim & Yanofsky (1980)
CGA	CTA	2.7	malT	−12	Chapon (1982)
*GAA	GTA	1.3	lac pr 115	−12	Maquat et al. (1980)
					Maquat & Reznikoff (1980)
GCA	GTA	2.8	dhu A	−12	Higgins & Ames (1982)
					Lee & Ames (1984)
GCT	GAT	1.6	lac P1	−16	Dickson et al. (1975)
					Reznikoff et al. (1982)
TAA	TTA	0	lpp	−12	Inouye & Inouye (1985)
TAC	TGC	−2.8	P22 ant	−14	Grana et al. (1985)
TCA	TAA	0.7	bio P98	−11	Otsuka & Abelson (1978)
TCA	TTA	0.7	Tn10 Pin	−11	Hawley & McClure (1983)
TGC	TAC	2.8	gal P2	−9	Busby et al. (1984)
*TGT	TAT	0.8	λ PI	−11	Abraham et al. (1980)
					Hoess et al. (1980)

Table I (Continued)

WT sequence	mutant sequence	free energy difference (kcal/mol)	description of mutation: gene	position	ref
			lac P1	−9	Gilbert (1976)
			araB	−10	Horwitz et al. (1980)
		−35 Region Down Mutations			
ACA	AAA	−0.6	lac P1	−32	Reznikoff et al. (1978)
					Reznikoff et al. (1982)
ACA	ATA	0.8	trp	−32	Yanofsky et al. (1981)
			λ PL	−31	Kleid et al. (1976)
			lac P1	−32	LeClerck & Istock (1982)
			P22 ant	−31	Youderian et al. (1982)
			Tn10 tetA	−31	Daniels & Bertrand (1985)
ACT	ATT	−0.5	λ PR	−31	Hawley & McClure (1983)
AGA	AAA	−0.6	λ PRM	−33	Meyer et al. (1975)
					Rosen et al. (1980)
AGG	ACG	−0.2	lac P1	−39	LeClerck & Istock (1982)
ATT	AAT	0	P22 ant	−35	Youderian et al. (1982)
ATT	AGT	0.5	P22 ant	−35	Youderian et al. (1982)
CAC	CGC	−3.5	lac P1	−31	Hawley & McClure (1983)
CGT	CAT	1.5	λ PRE	−35	Wulff et al. (1984)
CTT	CAT	0.1	lac P1	−36	LeClerck & Istock (1982)
CTT	CGT	−1.4	lac P1	−36	LeClerck & Istock (1982)
GAC	GCC	−3.3	P22 ant	−32	Youderian et al. (1982)
GAC	GGC	−3.3	λ PR	−32	Hawley & McClure (1983)
			P22 ant	−32	Youderian et al. (1982)
			Tn10 tetA	−32	Daniels & Bertrand (1985)
GAC	GTC	0	Tn10 tetA	−32	Daniels & Bertrand (1985)
GCG	GTG	3.5	λ PRE	−36	Wulff et al. (1980)
GCT	GTT	1.5	lac P1	−37	LeClerck & Istock (1982)
GGT	GAT	1.3	λ PRM	−39	Rosen et al. (1980)
GTA	GGA	−2.5	P22 ant	−37	Youderian et al. (1982)
GTG	GGG	−3.0	λ PRM	−38	Rosen et al. (1980)
GTT	GCT	−1.5	λ PRE	−39	Wulff et al. (1984)
			Tn10 tetA	−35	Daniels & Bertrand (1985)
GTT	GGT	−1.2	Tn10 tetA	−35	Daniels & Bertrand (1985)
TAT	TTT	−1.4	P22 ant	−36	Youderian et al. (1982)
TGA	TAA	0.7	P22 ant	−33	Youderian et al. (1982)
			R1 copA	−33	Stougaard et al. (1981)
			pMB1	−34	Cesareni (1982)
TGA	TTA	0.7	P22 ant	−33	Youderian et al. (1982)
TGC	TAC	2.8	λ PRE	−37	Wulff et al. (1980)
TGC	TTC	1.5	λ PRE	−37	Wulff et al. (1984)
TGT	TAT	0.8	bioB	−35	Barker et al. (1981)
TTA	TAA	0	lac P1	−34	LeClerck et al. (1984)
					Dickson et al. (1975)
TTA	TCA	−0.7	lac P1	−34	LeClerck & Istock (1982)
TTG	TCG	−1.4	Tn10 tetA	−34	Daniels & Bertrand (1985)
			P22 ant	−34	Youderian et al. (1982)
			λ PRE	−38	Wulff et al. (1984)
TTG	TGG	−1.2	λ PRE	−38	Wulff et al. (1984)
TTT	TCT	0.6	lac P1	−35	LeClerck & Istock (1982)
			λ PRE	−33	Wulff et al. (1980)
		−35 Region Up Mutations			
ATA	ACA	−0.8	λ PRM	−31	Meyer et al. (1980)
CGA	CAA	1.4	fol	−30	Smith et al. (1982)
CTC	CAC	0	lpp	−33	Inouye & Inouye (1985)
GAC	GGC	−3.3	araB	−35	Horwitz et al. (1980)
GCA	GTA	2.8	Tn10 Pout	−35	Hawley & McClure (1983)
GCG	GTG	3.5	lacI	−35	Calos (1978)
			rpoB	−34	An & Friesen (1980)
GTC	GAC	0	amp C	−32	Jaurin et al. (1982)
TCG	TTG	1.4	fol	−34	Smith et al. (1982)

[a] There are four sections based on the type of mutation (up/down) and on the location (−10 or −35 consensus regions). The first column lists the wild-type triplets in alphabetical order. The triplets including the mutation in the middle nucleotide are listed in the second column. The free energy difference is calculated by subtracting the wild-type triplet free energy from the mutant triplet free energy. The last three columns describe the mutations. The position of the mutation is indicated relative to the transcription start position (..., −5, −4, −3, −2, −1, →RNA, ...), and no gaps are introduced to align the consensus sequences. Asterisks indicate wild-type triplets for which a mutation in the middle nucleotide can either decrease or increase the free energy (see Discussion). [b] gal P2 mutations were isolated in a strain that contained a mutation G → A at position −14 of gal P1 (−9 of gal P2) [see Bingham et al. (1986)].

complex. (2) The −10 region is opened by thermal fluctuations. This step is fast and in equilibrium. (3) The enzyme interacts with the −10 region to form the open complex. Let RP_c^{-35} denote the closed complex at the −35 region, P_c^{-10} be the closed −10 region, P_o^{-10} be the open −10 region, and RP_o be the open complex. Then the reaction from closed to open complex can be described as

$$RP_c^{-35}P_c^{-10} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} RP_c^{-35}P_o^{-10} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} RP_o$$

Table II: Summary of Free Energy Comparisons between Promoter Mutations and the Corresponding Wild-Type Sequences[a]

	mutation type	increased free energy	decreased free energy	no change
−10 region	down	9	40	1
	up	14	2	2
−35 region	down	20	21	3
	up	5	2	2

[a] For each mutation type, we indicate the number of mutations that show increased free energy, decreased free energy, or no difference in comparison to the wild type.

where k_1 and k_{-1} are the rates of transition from helix to coil and from coil to helix, respectively, and k_2 and k_{-2} are the rates of the third step and its reverse, respectively. k_f is then

$$k_f = \frac{k_1 k_2}{k_{-1} + k_2}$$

If $k_{-1} \gg k_2$, then $k_f = (k_1/k_{-1})k_2$. The value k_1/k_{-1} is $e^{-\Delta G/RT}$ where ΔG is the free energy of transition from helix to coil in free solution. Hence, $k_f = e^{-\Delta G/RT}k_2$, and the overall rate is directly proportional to the equilibrium constant of opening near −10. [Note that the underlying assumption is that the available ΔG values (Breslauer et al., 1986) represent the situation in the presence of the enzyme. In other words, if the enzyme while interacting with the −35 region has any effect on the helix stability, this effect is not sequence dependent.] According to this model, the rate of open complex formation is directly proportional to the helix instability and to the rate of the last step.

This model predicts that, if k_2 is almost constant, then as the free energy of transition from helix to coil decreases (helix instability increases) k_f increases. We have used a database of 25 promoter sequences for which the rates of open complex formation have been measured (McClure, 1980; Dayton et al., 1984; Hawley & McClure, 1980, 1982; Stefano & Grala, 1982; Malan et al., 1984; Shih & Gussin, 1983; Simons et al., 1983; Mulligan et al., 1985) to calculate the correlation coefficient between the rate of open complex formation, k_f, and $e^{-\Delta G/RT}$ of short blocks within the promoter region. No significant correlation has been found in either the −35 or the −10 consensus regions. The lack of correlation cannot be due to the fact that $e^{-\Delta G/RT}$ is almost constant since there is a large variability in this value among the 25 promoters previously indicated to be 10^{-7}–10^{-10}. A similar study was carried out by Tachibana and Ishihama (1985) on a different set of 11 promoters. They characterized the relative rates of open complex formation by using a "mixed transcription" system (Kajitani & Ishihama, 1985; Nomura et al., 1985) and evaluated the melting properties of the DNA by using dinucleotide melting temperatures reported by Gotoh and Tagashira (1981b). They showed a high correlation ($r = 0.9$) between the opening potential of the double helix in the −9 to +3 region and the rate of open complex formation. However, when we use the free energies (Beslauer et al., 1986) on their database, we find a correlation of only $r = 0.3$ between the rate of open complex formation and $e^{-\Delta G/RT}$ for that segment. A similar result was also reported by Gotoh and Tagashira (1981b). For a numer of promoters, they found a very weak correlation between the promoter efficiency and the probability of base pairs opening. These analyses lead to the conclusion that if these are the steps of open complex formation then the rate is combined of two contributions and that the last rate-limiting step, characterized by k_2, is probably sequence dependent.

There are experimental measurements showing that in the transcription systems of T7 A1 (Chamberlin et al., 1982) and lac UV5 (Buc & McClure, 1985) the DNA melting is not a rate-limiting step in the open complex formation. Accordingly, it has been suggested (Chamberlin et al., 1982; Buc & McClure, 1985) that the open complex formation consists of a rate-limiting intermediate step followed by DNA melting. For this model, the reaction from closed to open complex can be described as

$$RP_c \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} RP_i \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} RP_o$$

where RP_i is the intermediate complex, k_1 and k_{-1} are, as before, the rates of transition from helix to coil and from coil to helix, respectively, and k_2 and k_{-2} are the rates of the intermediate step and its reverse, respectively. k_f is then

$$\frac{k_2 k_1}{k_{-2} + k_1}$$

and k_r is

$$\frac{k_{-2} k_{-1}}{k_{-2} + k_1}$$

If $k_1 \gg k_{-2}$, then $k_f = k_2$ and $k_r = (k_{-1}/k_1)k_{-2}$. Hence, in this model, k_f depends only on the rate of the intermediate complex formation and not on the helix stability, while k_r depends on the free energy of transition from coil to helix and on the rate of the reverse intermediate process (k_{-2}). However, when we use the k_r values available for 10 promoters (McClure, 1980; Dayton et al., 1984; Hawley & McClure, 1980; Malan et al., 1984; Mulligan et al., 1985) to calculate the correlation with the free energies, again, no correlation is found. It should be noted that the measured kinetic values may have large experimental errors and include measurements at different salt concentrations. Hence, a larger set of k_f and k_r values measured under constant conditions would afford more reliable conclusions on the relationship between the double-helix stability and the process of open complex formation. However, according to these considerations, it is hard to confirm any of the models. A possible conclusion, relevant to both models, is that while there is a general tendency of the −10 regions to be relatively unstable, this property by itself cannot explain the variation in isomerization rates among individual promoters.

Several findings question the validity of a strictly bipartite model in which the −35 region is responsible only for the recognition and the −10 region is responsible for the open complex formation. Mutations affecting the isomerization rate are found in both the −10 and −35 regions (Stefano & Gralla, 1982; McClure et al., 1982). Also, in our study, 20% of the mutations in the −10 region do not show a correlation between the direction of free energy change and the type of mutation. It is conceivable that the isomerization to open complex depends on an appropriate positioning of the RNA polymerase and on specific contacts in the −10 region as well as on its ease of melting. Hence, mutations in the −35 region which affect the isomerization rate may hinder the optimal positioning of the enzyme, and the 20% of the mutations mentioned above, which do not behave as expected, may affect the specific contacts with the enzyme rather than the double-helix stability. If this were the case, then we would not expect to find a correlation with the kinetic data. The database for which kinetic data are available consists of sequences with differences in both the −10 and −35 consensus regions, and the variations in their k_f and k_r values might be due to other sequence-de-

pendent factors in addition to helix stability. However, the situation for the mutations is less complex. We are performing a *pairwise* comparison between a mutant and a wild type. The two sequences are identical except in one position. Hence, all the sequence-dependent factors that could affect the open complex formation are identical except the one dependent on the position of the mutation. Our analysis shows that 80% of those up/down mutations in the −10 region are correlated with increased/decreased free energy. this clearly suggests that the opening transition has an effect on the overall transcription rate. The highly significant correlation between the type of mutation and the direction of free energy change, as well as the strong peak in the free energy distribution, indicates that instability at the −10 region plays a role in the steps preceding transcription initiation.

Acknowledgments

We thank Dr. Jacob Maizel for discussions and for his continued interest and support. We are grateful to Drs. Donald Court, David Covell, Hans Westerhoff, and Sanzo Miyazawa for their helpful suggestions and comments.

Supplementary Material Available

A list of the 168 *E. coli* sequences used for the analysis of relative helix stability of the promoter region (7 pages). Ordering information is given on any current masthead page.

References

Abraham, J., Mascarenhas, D., Fischer, R., Benedik, M., Campbell, A., & Echols, H. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2477–2481.

An, G., & Friesen, J. D. (1980) *J. Bacteriol. 144*, 904–916.

Barker, D. F., Kuhn, J., & Campbell, A. M. (1981) *Gene 13*, 89–102.

Berman, M. L., & Landy, A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4303–4307.

Bingham, A. H. A., Ponnambalam, S., Chan, B., & Busby, S. (1986) *Gene 41*, 67–74.

Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–3750.

Buc, H., & McClure, W. R. (1985) *Biochemistry 24*, 2712–2723.

Busby, S., Aiba, H., & de Crombrugghe, B. (1982) *J. Mol. Biol. 154*, 211–227.

Busby, S., Truelle, N., Spassky, A., Dreyfus, M., & Buc, H. (1984) *Gene 28*, 201–209.

Calos, M. P. (1978) *Nature (London) 274*, 762–765.

Cesareni, G. (1982) *J. Mol. Biol. 160*, 123–126.

Chamberlin, M. J., Rosenberg, S., & Kadesch, T. (1982) in *Promoters: Structure and Function* (Rodriguez, R. L., & Chamberlin, M. J., Eds.) pp 34–53, Praeger, New York.

Chapon, C. (1982) *EMBO J. 1*, 369–374.

Daniels, D. W., & Bertrand, K. P. (1985) *J. Mol. Biol. 184*, 599–610.

Dayton, C. J., Prosen, D. E., Parker, K. L., & Cech, C. L. (1984) *J. Biol. Chem. 259*, 1616–1621.

Dickson, R. C., Abelson, J., Barnes, W. M., & Reznikoff, W. S. (1975) *Science (Washington, D.C.) 187*, 27–35.

Gemmill, R. M., Tripp, M., Friedman, S. B., & Calvo, J. M. (1984) *J. Bacteriol. 158*, 948–953.

Gilbert, W. (1976) in *RNA Polymerase* (Losick, R., & Chamberlin, M., Eds.) pp 193–205, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Gotoh, O., & Tagashira, Y. (1981a) *Biopolymers 20*, 1033–1042.

Gotoh, O., & Tagashira, Y. (1981b) *Biopolymers 20*, 1043–1058.

Grana, D., Youderian, P., & Susskind, M. M. (1985) *Genetics 111*, 1–16.

Harley, C. B., & Reynolds, R. P. (1987) *Nucleic Acids Res. 15*, 2343–2361.

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.

Higgins, C. F., & Ames, G. F.-L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1083–1087.

Hoess, R. H., Foeller, C., Bidwell, K., & Landy, A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2482–2486.

Horwitz, A. H., Morandi, C., & Wilcox, G. (1980) *J. Bacteriol. 142*, 659–667.

Inouye, S., & Inouye, M. (1985) *Nucleic Acids Res. 13*, 3101–3110.

Jaurin, B., Grundstrom, T., & Normark, S. (1982) *EMBO J. 1*, 875–881.

Kajitani, M., & Ishihama, A. (1983) *Nucleic Acids Res. 11*, 671–686.

Kleid, D., Humayun, Z., Jeffrey, A., & Ptashne, M. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 293–297.

LeClerc, J. E., & Istock, N. L. (1982) *Nature (London) 297*, 596–598.

LeClerc, J. E., Istock, N. L., Saran, B. R., & Allen, R., Jr. (1984) *J. Mol. Biol. 180*, 217–237.

Lee, G. S., & Ames, G. F.-L. (1984) *J. Bacteriol. 159*, 1000–1005.

Malan, T. P., Kolb, A., Buc, H., & McClure, W. R. (1984) *J. Mol. Biol. 180*, 881–909.

Maquat, L. E., & Reznikoff, W. S. (1980) *J. Mol. Biol. 139*, 551–556.

Maquat, L. E., Thornton, K., & Reznikoff, W. S. (1980) *J. Mol. Biol. 139*, 537–549.

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

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

McClure, W. R., Hawley, D. K., & Malan, T. P. (1982) in *Promoters: Structue and Function* (Rodriguez, R. L., & Chamberlin, M. J., Eds.) pp 111–120, Praeger, New York.

Meyer, B. J., Kleid, D. G., & Ptashne, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4785–4789.

Meyer, B. J., Maurer, R., & Ptashne, M. (1980) *J. Mol. Biol. 139*, 163–194.

Miozzari, G., & Yanofsky, C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5580–5584.

Mozola, M. A., Friedman, D. I., Crawford, C. L., Wulff, D. L., Shimatake, H., & Rosenberg, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1122–1125.

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

Musso, R. E., DiLauro, R., Adhya, S., & deCrombrugghe, B. (1977) *Cell (Cambridge, Mass.) 12*, 847–854.

Nakanishi, S., Adhya, S., Gottesman, M., & Pastan, I. (1974) *J. Biol. Chem. 249*, 4050–4056.

Nomura, T., Nobuyuki, F., & Ishihama, A. (1985) *Nucleic Acids Res. 13*, 7647–7661.

Nussinov, R. (1984) *J. Biol. Chem. 259*, 6798–6805.

Nussinov, R. (1985) *J. Theor. Biol. 115*, 179–189.
Oppenheim, D. S., & Yanofsky, C. (1980) *J. Mol. Biol. 144*, 143–161.
Oppenheim, D. S., Bennett, G. N., & Yanofsky, C. (1980) *J. Mol. Biol. 144*, 133–142.
Otsuka, A., & Abelson, J. (1978) *Nature (London) 276*, 689–694.
Piette, J., Cunin, R., Boyen, A., Charlier, D., Crabeel, M., Van Vliet, F., Glansdorff, N., Squires, C., & Squires, C. L. (1982) *Nucleic Acids Res. 10*, 8031–8048.
Post, L. E., Arfsten, A. E., Nomura, M., & Jaskunas, S. R. (1978) *Cell (Cambridge, Mass.) 15*, 231–236.
Reznikoff, W. S., & Abelson, J. N. (1978) in *The Operon* (Miller, J. H., & Reznikoff, W. S., Eds.) pp 221–243, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Reznikoff, W. S., Maquat, L. E., Munson, L. M., Johnson, R. C., & Mandecki, W. (1982) in *Promoters: Structure and Function* (Rodriguez, R. L., & Chamberlin, M. J., Eds.) pp 80–95, Praeger, New York.
Rosen, E. D., Hartley, J. L., Matz, K., Nichols, B. P., Young, K. M., Donelson, J. E., & Gussin, G. N. (1980) *Gene 11*, 197–205.
Rosenberg, M., Court, D., Shimatake, H., Brady, C., & Wulff, D. L. (1978) *Nature (London) 272*, 414–423.
Rothstein, S. J., & Reznikoff, W. S. (1981) *Cell (Cambridge, Mass.) 23*, 191–199.
Shih, M. C., & Gussin, G. N. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 496–500.
Siebenlist, U., Simpson, R. B., & Gilbert, W. (1980) *Cell (Cambridge, Mass.) 20*, 269–281.
Simons, R. W., Hoopes, B. C., McClure, W. R., & Kleckner, N. (1983) *Cell (Cambridge, Mass.) 34*, 673–682.
Smith, D. R., Rood, J. I., Bird, P. I., Sneddon, M. K., Calvo, J. M., & Morisson, J. F. (1982) *J. Biol. Chem. 257*, 9043–9048.
Stefano, J. E., & Gralla, J. D. (1982) *J. Biol. Chem. 257*, 13924–13929.
Stougaard, P., Molin, S., & Nordstrom, K. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6008–6012.
Tachibana, H., & Ishihama, A. (1985) *Nucleic Acids Res. 13*, 9031–9042.
von Hippel, P. H., Bear, D. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.
Wang, J. C. (1982) in *Promoters: Structure and Function* (Rodiguez, R. L., & Chamberlin, M. J., Eds.) pp 229–241, Praeger, New York.
Wilcox, G., Al-Zarban, S., Cass, L. G., Clarke, P., Heffernan, L., Horwitz, A. H., & Miyada, C. G. (1982) in *Promoters: Structure and Function (Rodriguez*, R. L., & Chamberlin, M. J., Eds.) pp 183–194, Praeger, New York.
Wulff, D. L., Beher, M., Izumi, S., Beck, J., Mahoney, M., Shimatake, H., Brady, C., Court, D., & Rosenberg, M. (1980) *J. Mol. Biol. 138*, 209–230.
Wulff, D. L., Mahoney, M., Shatzman, A., & Rosenberg, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 555–559.
Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., VanCleemput, M., & Wu, A. M. (1981) *Nucleic Acids Res. 9*, 6647–6668.
Youderian, P., Bouvier, S., & Susskind, M. M. (1982) *Cell (Cambridge, Mass.) 30*, 843–853.

Isolation of Multiple Forms of DNA Polymerase δ: Evidence of Proteolytic Modification during Isolation†

Marietta Y. W. T. Lee*
Department of Medicine, University of Miami School of Medicine, Miami, Florida 33101
Received April 14, 1987; Revised Manuscript Received March 14, 1988

ABSTRACT: The subunit structures of a number of human placenta DNA polymerase δ preparations were investigated by Western blotting with polyclonal antisera and by activity staining following polyacrylamide gel electrophoresis. When immunoblots and activity stains were performed on different enzyme preparations, putative catalytic subunits of (a) 170, (b) 120, or (c) 50–70 kilodaltons (kDa) were observed. It was also observed that the lower molecular weight forms could be generated upon storage of the preparations. Western blotting of human placental tissue extracts showed that the major immunoreactive polypeptide was 160–170 kDa. Treatment of the extracts with trypsin or *Staphylococcus aureus* V8 protease led to the generation of immunoreactive polypeptides of lower molecular weights. These studies suggest that the 120-kDa and lower forms of the enzyme are generated via uncontrolled proteolysis and provide a rationale for the observation of different apparent subunit structures previously reported for DNA polymerase δ. In addition, these findings suggest that DNA polymerase δ has a catalytic domain which resides in a protease-resistant domain.

DNA polymerase δ is a high molecular weight DNA polymerase which has been described in rabbit reticulocytes (Byrnes, 1984), calf thymus (Lee et al., 1981, 1984; Crute et al., 1986), and human placenta (Lee & Toomey, 1987). DNA polymerase δ is of particular interest since it differs from DNA polymerase α in that it is associated with a 3′ to 5′ exonuclease activity which is not removed by any conventional fractionation methods, and therefore appears to be intrinsic to the enzyme. We have reported studies of a highly purified DNA polymerase δ preparation from human placenta which possessed a 170-kilodalton (kDa) polypeptide, with smaller polypeptides in the 60–70-kDa range; in this study, murine antisera were shown

† This work was supported by NIH Grant GM 31973 and was performed during the tenure of an Established Investigatorship of the American Heart Association to M.Y.W.T.L.
* Address correspondence to this author at the Department of Medicine (R-57), University of Miami School of Medicine, P.O. Box 016960, Miami, FL 33101.

0006-2960/88/0427-5188$01.50/0