Volume 4 Number 8 August 1977 Nucleic Acids Research

The transmission of stability or instability from site specific protein-DNA complexes

Roger M. Wartell

School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received 17 May 1977

ABSTRACT

Theoretical calculations were made to determine the influence of site specific 'melting' and 'stabilizing' proteins on the thermal stability of nearby base pairs (bp.). A DNA sequence 999bp. long containing the 123 bp. lactose operon control region in the center was examined. Melting curves of base pairs near the binding sites of the catabolite activator protein, CAP, the lactose repressor, and RNA polymerase were calculated in the absence and presence of each protein. The empirical loop entropy model of the helix-coil transition of DNA was employed. Calculations show that melting and stabilizing proteins alter the t_m of base pairs 20 to 100 bp. away. The magnitude and range of the effect is strongly influenced by the base pair composition and sequence of the protein site and the immediately adjacent DNA regions.

INTRODUCTION

Over the past two decades much evidence has shown that transcription is controlled by the interaction of RNA polymerase and affector proteins (repressors and activators) with specific DNA regions. Site specific protein-DNA interaction is the molecular basis for other metabolic processes associated with DNA as well. A complete understanding of the highly specific recognition process is not available. A basic feature of DNA which must be involved in protein recognition is the position in space of specific chemical groups. The nucleotide sequence of the site will obviously determine what chemical groups are available. Their positions will be determined by the conformations of the site. Due to cooperative interactions along DNA, and the dynamic character of DNA in solution, the regions surrounding the protein site may help determine the positions of the interactive groups. An extension of this idea is that a 'director protein' binding to one specific DNA site may alter a nearby DNA region. This could enhance or decrease the binding of a second protein.[1,2] This type of model has been proposed to explain how the catabolite activator protein (CAP) enhances RNA polymerase binding.[2,3] Although several studies have been made to examine the influence of one DNA region on adjacent ones,[2,4,5] no quantitative estimates exist on the local influence of a site specific protein.

In this work I examined the influence of site specific 'melting' and 'stabilizing' proteins on the thermal stability of base pairs 20 to 100 bp. away from the site. This was done by calculating the probability that a base pair is hydrogen bonded in the presence and absence of a nearby site specific protein. The transmission of stability or instability (telestability) to the base pair was quantified by the difference in its melting temperatures with and without protein.

The loop entropy model of the helix coil transition of DNA was employed. This model is an empirical one. It has been used to determine the thermodynamic properties of nucleic acid oligomers and polymers.[6-8] The DNA sequence examined is the Escherichia coli lactose operon control region (Figure 1).

Although this study cannot determine detailed information on conformational changes, it provides estimates of transmitted free energy changes. The results show that melting and stabilizing proteins alter the stability of the DNA regions surrounding them. Local nucleotide composition and sequence strongly influence the magnitude and range of the transmitted changes.

THEORETICAL METHODS

A. Melting of a base pair inside a long DNA

The loop entropy model of the DNA helix-coil transition has been previously described.[8,9,10] Several melting curve calculations have been presented using this model for a DNA with a random distribution of A·T and G·C base pairs.[8-12] Crothers showed that the probability of melting one internal base pair strongly depends on the local base pair composition. This calculation employed a coarse

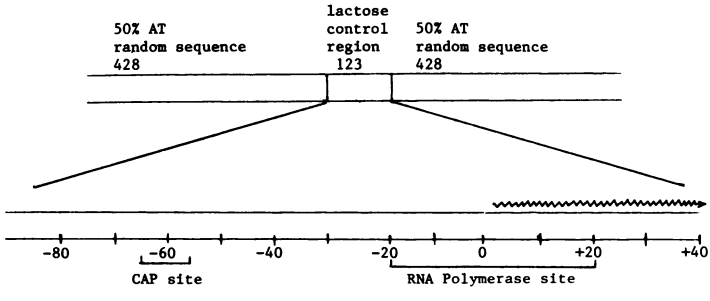


Figure 1: The 999 base pair DNA sequence which was examined. The lactose operon region is shown with the positions of the CAP site, RNA polymerase site and the initiation of mRNA (~~~). The nucleotide sequence was taken from reference 3. Numbering of the lactose operon region is in base pairs with +1 being the first base pair transcribed into mRNA.

graining approximation. 'Coarse graining' groups base pairs together into blocks and treats each block as a unit. A more recent calculation[8] showed that when coarse graining is removed, the above conclusion still holds. However, coarse graining results in substantial error with respect to the quantitative influence of local base pair composition on melting behavior.[8,12] In the calculation described below computational approximations are removed. Assumptions of the model are discussed and their influence on the results examined.

1. Basic theory

Each base pair of the DNA is considered to be in one of two states, hydrogen bonded (intact) or non hydrogen bonded (broken). The free energy change in forming the i^{th} base pair depends on the type of base pair (A·T or G·C) and on the type and condition of nearest neighbor base pairs. An equilibrium constant $\underline{s}$ is assigned to the reaction of forming an intact base pair next to one intact pair. When an isolated base pair closes an internal loop of m base pairs one assigns the equilibrium constant $\sigma\underline{s}f(m)$ to the reaction. σ removes the stacking energy from $\underline{s}$, since base pair stacking does not occur in this reaction. A long range effect is included by assigning the equilibrium constant f(m) for forming a loop of m melted base pairs sandwiched between duplex sections. f(m) accounts for the configurational entropy of the strands of an internal loop of m base pairs relative to an equivalent number of unbonded base pairs at an open end. Formation of the first isolated base pair joining the two single stranded DNAs is denoted by the equilibrium constant $\beta\underline{s}$.

It is recognized that the assumption that only nearest neighbor interactions contribute to $\underline{s}$ is a formal one. In reality, longer range coulombic forces are averaged into the equilibrium constants. The average lengths of duplex regions are sufficiently long through most of the transition that nearly all intact base pairs experience a constant coulombic environment. A similar argument holds for the unbonded regions. Providing that the longer range forces are not both strong and highly base pair dependent, one can average them into a nearest neighbor interaction for most base pairs. Thus the nearest neighbor assumption should not influence calculated transition behavior when bonded and unbonded regions are long and statistically distributed. What is not considered in this calculation is how a melted out or thermally stabilized region localized to a specific site influences adjacent regions through interactions extending beyond nearest nucleotide neighbors. Since this work concerns the region surrounding such a site, this is a significant consideration. Recent analysis of the melting curves of A·T/G·C block DNA oligomers in the presence of actinomycin showed that interactions extending beyond neighboring nucleotides can be signifi-

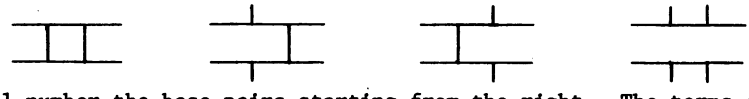
cant.[5] The nature of these interactions is not clear. Thus the energy changes calculated in this work should not be overinterpreted. The qualitative conclusions on how base pair sequence influences telestability should be generally valid.

2. Computational method

The method of calculating the DNA partition function and the probability of base pair melting is similar to one previously described in detail.[8] The small changes in procedure employed here provide a more accurate correlation between the DNA configurations and their statistical weights. Given a specific sequence of N base pairs with N_A A·T and N_G G·C pairs, we separate the partition function Z_N into four groupings:

1. Z_N^{hh}. All configurations starting and ending with a bonded or 'helical' base pair.
2. Z_N^{hc}. All configurations starting with a helical base pair and ending with an unbonded or 'coil' base pair.
3. Z_N^{ch}. All configurations starting with a coil pair and ending with a helical pair.
4. Z_N^{cc}. All configurations starting and ending with a coil base pair.

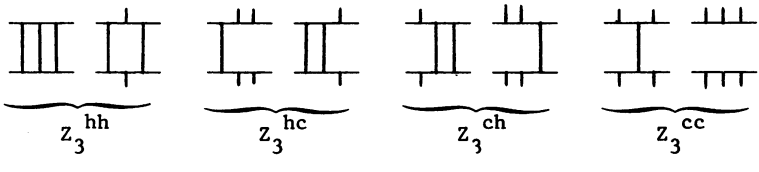
The partition function is the sum of these four terms. The method of calculating Z_N is obtained by induction. The cases N=2, N=3 will be shown in detail. For the case N=2, the possible configurations sketched in the 'ladder representation' are



We will number the base pairs starting from the right. The terms described above are

$$Z_2^{hh} = \beta s_{\underline{1}} s_{\underline{2}} \qquad Z_2^{hc} = \beta s_{\underline{2}}$$
$$Z_2^{ch}(1) = \beta s_{\underline{1}} \qquad Z_2^{cc}(2) = 1 \qquad (1)$$

$Z_n^{ch}(i)$ denores the configurations of Z_n^{ch} which have i broken bonds looking in from the left. A similar definition holds for $Z_n^{cc}(i)$ with respect to Z_n^{cc}. For N=3 one has the configurations



where

$$Z_3^{hh} = \beta\underline{s}_1\underline{s}_2\underline{s}_3 + \beta\underline{s}_3\sigma f(1)\underline{s}_1, \quad Z_3^{ch}(1) = \beta\underline{s}_1\underline{s}_2, \quad Z_3^{ch}(2) = \beta\underline{s}_1$$

$$Z_3^{hc} = \beta\underline{s}_3 + \beta\underline{s}_3\underline{s}_2, \quad Z_3^{cc}(1) = \beta\underline{s}_2, \quad Z_3^{cc}(3) = 1 \qquad (2)$$

Z_3 can be written as

$$Z_3 = (1 \;\; 1) \begin{pmatrix} \underline{s}_3 & \underline{s}_3\underline{0} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} Z_2^{hh} & Z_2^{hc} \\ \sum_{i=1}^{1} Z_2^{ch}(i) & \sum_{i=1}^{2} Z_2^{cc}(i) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad (3)$$

with $\underline{0}$ an operator with the properties

$$\underline{0} \sum_{i=1}^{1} Z_2^{ch}(i) = \sigma \sum_{i=1}^{1} f(i)\, Z_2^{ch}(i) \qquad (4a)$$

$$\underline{0} \sum_{i=1}^{2} Z_2^{cc}(i) = \sigma \sum_{i=1}^{1} f(i)\, Z_2^{cc}(i) + \beta\, Z_2^{cc}(2) \qquad (4b)$$

Continuing the process to higher N one finds

$$Z_N = G_N + H_N = (1 \;\; 1) \prod_{n=3}^{N} \begin{pmatrix} \underline{s}_n & \underline{s}_n\underline{0} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} G_{n-1} \\ H_{n-1} \end{pmatrix} \qquad (5)$$

with

$$G_n = Z_n^{hh} + Z_n^{hc}, \quad H_n = \sum_{i=1}^{n-1} Z_n^{ch}(i) + \sum_{i=1}^{n} Z_n^{cc}(i) \qquad (6)$$

The process used to obtain Z_3 above is in general for Z_n

$$\underline{0} \sum_{i=1}^{n-2} Z_{n-1}^{ch}(i) = \sigma \sum_{i=1}^{n-2} f(i)\, Z_{n-1}^{ch}(i) \qquad (7a)$$

$$\underline{0} \sum_{i=1}^{n-1} Z_{n-1}^{cc}(1) = \sum_{i=1}^{n-2} f(i)\, Z_{n-1}^{cc}(i) + \beta\, Z_{n-1}^{cc}(n-1) \qquad (7b)$$

with

$$\left.\begin{aligned} Z_n^{ch}(1) &= Z_{n-1}^{hh}; & Z_n^{ch}(i) &= Z_{n-1}^{ch}(i-1) \\ Z_n^{cc}(1) &= Z_{n-1}^{hc}; & Z_n^{cc}(i) &= Z_{n-1}^{cc}(i-1) \end{aligned}\right\} \quad i>1 \qquad (7c)$$

To obtain the probability that the j^{th} base pair is broken one needs

$$\theta_j(T) = 1 - \frac{\partial_j G_N + \partial_j H_N}{G_N + H_N} \qquad (8a)$$

where $\partial_j = \underline{s}_j \frac{\partial}{\partial \underline{s}_j}$ (8b)

One method of calculating $\theta_j(T)$ is to calculate Z_N from equation (5) and then obtain $(\partial_j G_N + \partial_j H_N)$ from equation (5) by replacing the matrix term

$$\begin{pmatrix} \underline{s}_j & \underline{s}_j & \underline{0} \\ 1 & 1 & \end{pmatrix} \quad \text{with} \quad \begin{pmatrix} \underline{s}_j & \underline{s}_j & \underline{0} \\ 0 & 0 & \end{pmatrix}$$

This change removes all the terms in Z_N where $\underline{s}_j$ doesn't appear.

3. Selection of parameters

Values of the thermodynamics parameters were chosen to correspond whenever possible to 0.2 M sodium ion. Equilibrium association constants $\underline{s}_{AT}$ and $\underline{s}_{GC}$ were assigned to A·T and G·C pairs respectively. The dependence of these constants on the chemical character of their neighboring base pairs was omitted. This effect can be included if it is desired at a future time. The entropy change for forming a stacked base pair was taken to be the same for A·T and G·C pairs, $\Delta S = -24.8$ ev./mole base pairs. The enthalpy change in this process was -8.5 kcal/mole base pair for A·T, and -9.4 kcal/mole base pairs for G·C and assumed to be independent of temperature. The entropy value was obtained from calorimetric measurements by Scheffler and Sturtevant.[13] The enthalpy values was determined from the extrapolated melting temperatures for A·T and G·C pairs in naturally occurring DNAs in 0.2 M sodium ion. The melting temperature of an A·T pair was 68.°C and for a G·C pair, 112°C. The temperature dependence of $\underline{s}_{AT}$ and $\underline{s}_{GC}$ are given by the Boltzman relation of equilibrium constants and free energy. The stacking energy term employed, $\sigma = 4.5 \times 10^{-5}$, was estimated from values measured for A·T and G·C DNA polymer transitions.[14,15] Changes in this value by ±50% did not greatly alter the results. The loop entropy function was empirically chosen to provide the best fit to the loop free energies evaluated by Gralla and Crothers[16] for small loops of RNA and the loop weighting function for large DNA loops.[14,15] The function was

$$f(m) = 0.006/((1 - e^{-.049m})(m + 1))^{1.55} \qquad (9)$$

This functional form is the best available estimate. Using the above parameters and methods, melting curves were calculated for specific base pairs inside a 999 base pair long DNA.

B. Relating a protein's binding properties to the thermal stability of the DNA site

This section will relate the equilibrium binding constant and site size of a protein-DNA complex to the thermal stability of the site. A number of studies have examined the influence of ligand binding on the helix coil transition of DNA.[17-19] The equations presented are similar to these earlier works, however the emphasis is on local effects. Two types of binding proteins will be examined; one which stabilizes the duplex form of DNA, the 'stabilizing protein' and one which forms a melted out region of DNA in its final complex, the 'melting protein'. It is assumed the protein binds to only one specific site on the piece of DNA being examined. The effect of simultaneous but weaker binding to independent non-specific sites elsewhere is also considered.

The four possible initial and final states being considered for the protein-DNA interaction at the specific site are sketched in Figure 2. The $K_i (i = 1 - 4)$ are the association constants for the reactions. The first and last reactions are for a stabilizing protein, the second and third are for a melting protein. We denote by d the concentration of empty sites in the duplex form and by b the concentration of empty sites in the bubble or melted form. P is the free protein concentration and D and B represent the complexes of DNA with protein. Clearly a large number of steps can occur between the initial and final states shown above. One point worth noting is that a melting protein need not recognize a thermally liable region, i.e. b. Around 25 - 40°C one can expect d >> b. Therefore a rapidly binding protein which is observed in a melted complex B will most likely recognize its DNA site as a duplex. We assume this to be the case below.

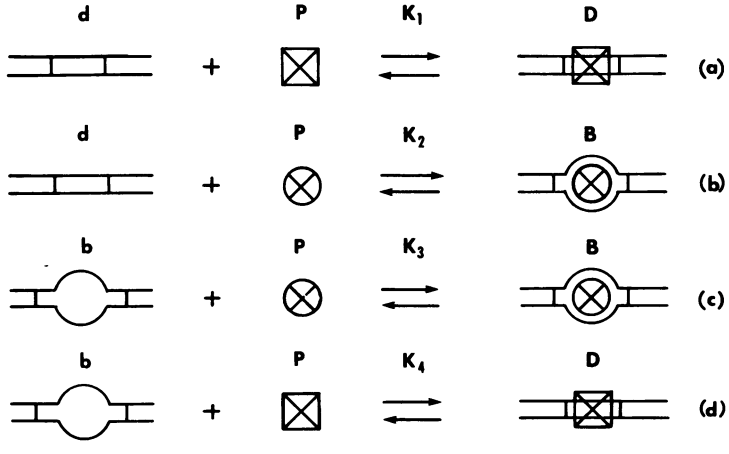


Figure 2: The initial and final states of four protein-DNA interactions considered in this work. The square objects represent 'stabilizing proteins,' the round objects represent 'melting proteins.'

The binding of a melting protein to a specific site is examined first. The protein is assumed to melt out m base pairs. K_1 and K_4 will be set equal to zero. It is also assumed the $K_3 = 0$. Possible examples of such a protein are RNA polymerase and the cAMP · CAP protein complex. When no protein is present, the relative probability of the site being duplex/melted is

$$R_o = d/b = \underline{s}_{AT}^{m_a} \underline{s}_{GC}^{m_g}/\sigma f(m)$$

$$= \frac{\exp[-(m_a \Delta G_{AT} + m_g \Delta G_{GC})/RT]}{\exp[-(\Delta G(m) - \Delta G_s)/RT]} \qquad (10)$$

m_a and m_g are the number of A·T and G·C pairs at the site ($m_a + m_g = m$). ΔG_{AT} and ΔG_{GC} are the free energies of base pair formation. $\Delta G(m)$ is the loop entropy free energy and ΔG_s the stacking energy per base pair. Equation (10) can be deduced by examining the statistical weights of short duplex DNAs, eg. N=4, with and without an internal loop. Writing ΔG_{AT} and ΔG_{GC} in terms of enthalpy and entropy, and employing the assumption that the entropy, ΔS, is the same for both base pairs, R_o is

$$R_o = \frac{\exp[(\Delta S/RT)(m_a(T-T_A) + m_g(T-T_G))]}{f(m)\sigma} \qquad (11)$$

T is the temperature at which one is examining the melting curve. T_A and T_G are the temperatures where ΔG_{AT} and ΔG_{GC} equal zero.

When the melting protein is present, the probability that the site is duplex/melted is

$$R_p = d/(b + B) = d/(b + K_2 P d) = R_o(1 + K_2 P R_o)^{-1} \qquad (12)$$

For K_2P greater than one, K_2PR_o is much greater than one for temperatures in and below the T_m region of the DNA. Since this is usually the case, a very good approximation for R_p in and below the transition region is

$$R_p = (K_2 P)^{-1} \qquad (13)$$

R_p will be empirically expressed in terms of new equilibrium constants $\underline{s}^*_{AT}$ and $\underline{s}^*_{GC}$ for the protein site. This will allow the use of the DNA melting theory presented above.

$$R_p = (K_2 P)^{-1} = (\underline{s}^*_{AT})^{m_a} (\underline{s}^*_{GC})^{m_g}/(f(m)\sigma)$$

$$= \exp[\Delta S/RT(m_a(T-T^*_A) + m_g(T-T_G*))]/(f(m)\sigma) \qquad (14)$$

Letting

$$T_A^* = T_A - \Delta T$$

$$T_G^* = T_G - \Delta T$$

one obtains

$$\Delta T = (-RT/m\Delta S)\ell n[K_2 P R_o] \quad (15)$$

When melting proteins are present, one lowers T_A and T_G of base pairs in the protein site by ΔT. When the melting protein binds to independent non-specific sites in addition to the specific site, K_2 is replaced by $K_2/(1 + K_N S)$ where K_N is the association constant for nonspecific binding and S is the total concentration of non-specific sites.[20] Here it is assumed that the non-specific binding occurs by the reaction sketched in Figure 2a.

The binding of a stabilizing protein to a specific site is described by Figure 2a. The lactose repressor and λ phage cI gene repressor are thought to be examples of this reaction. The probability that the specific site is duplex/melted is

$$R_p' = (d + D)/b = d(1 + K_1 P)/b$$

$$= R_o(1 + K_1 P) \quad (16)$$

Following equations (10) to (15) R_p' can be expressed in terms of new binding constants s_{AT}' and s_{GC}' for the base pairs at the site. T_A and T_G of these base pairs are <u>increased</u> by $\Delta T'$ where

$$\Delta T' = -(RT/m\Delta S)\ell n(1 + K_1 P) \quad (17)$$

When the stabilizing protein also forms complexes at independent non-specific sites in addition to the specific site, K_1 is replaced by $K_1/(1 + K_N S)$. Non-specific binding is also assumed to occur via Figure 2a.

RESULTS

Melting curves for specific base pairs inside the 999 base pair DNA shown in Figure 1 were obtained. This DNA has a randomly generated A·T/G·C sequence for 438 base pairs before and after the lac sequence. The base pair sequence -83 to +38 is the lactose operon control region, i.e. between the end of the i gene and the beginning of the β-galactosidase gene. The <u>a priori</u> probability of an A·T pair for positions in the random region was 50%. The same random sequence is generated for all calculations, and the <u>a posteriori</u> % A·T content for the random regions was 50.6. The DNA is long enough so that end results

do not influence the central 150 base pairs. Changing ten base pairs at the ends to G·C pairs did not alter the results.

The melting temperature, t_m, is the temperature where the probability of a base pair being broken, $\theta(t)$, equals 0.5 (t is temperature in ^oC, T is temperature in ^oK). Equation (15) or equation (17) are used to account for the protein's influence on the stability parameters of its site. The temperature, T, in the above equations must be the temperature at which the melting curve is being calculated. The rigorously correct approach would be to calculate the probability of base pair melting around 25-40^oC, where the protein-DNA binding constants are measured. In this temperature range, however, the equilibrium probability of melting is small. To avoid the difficulty of interpreting differences between small numbers far from the transition, the protein -DNA binding constants are assumed to be independent of temperature.

A. Influence of site specific protein on surrounding base pairs

1. Melting protein effect

The effect of a melting protein binding to the CAP site was examined. This simulates a previously suggested model for explaining how the CAP·cAMP factor when bound to DNA enhances RNA polymerase action.[2,3] The CAP site was taken to be 14 base pairs centered between base pairs -60 and -61. Ten base pairs, -65 to -56 were assumed to melt out due to CAP binding. The melting curves of base pairs between -47 and +38 were examined with and without melting protein at the CAP site. Wu et al[21] has determined that the association constant of CAP protein to DNA is much greater than $10^9 M^{-1}$ in a buffer of ionic strength 0.22. A binding constant of $K_2 = 10^{12} M^{-1}$ was assumed for this work. Changes in K_2 by an order of magnitude do not alter the t_m changes by more than ±.02^oC. This indicates the results are not too sensitive to the K_2 value. Also, the binding constant was effectively lowered by including non-specific binding. A cellular concentration for CAP protein was estimated from experimental data to be $5 \times 10^{-6} M$.[22] The non-specific binding constant was taken to be $K_N = 10^{+6} M^{-1}$ and the non-specific DNA concentration was $10^{-4} M$. The latter value is close to the total DNA concentration while the former is the association constant estimated for non-specific binding of other site specific proteins.

Figure 3 shows the melting curves of base pair -37 of the lac sequence. The curve on the right shows the transition in the absence of protein. The curve on the left shows the transition with a melting protein located at the CAP site. The melting protein lowers the transition midpoint of this base pair by 0.92^oC and broadens the transition. A similar behavior was observed for other base pairs examined. Figure 4 shows the influence of a melting protein at the CAP

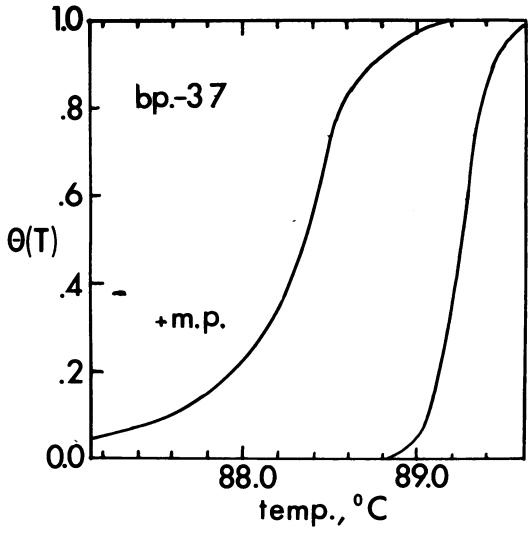


Figure 3: Melting curves of base pair -37 in the absence of any protein (curve on right) and with a melting protein at the CAP site (+ m.p.)

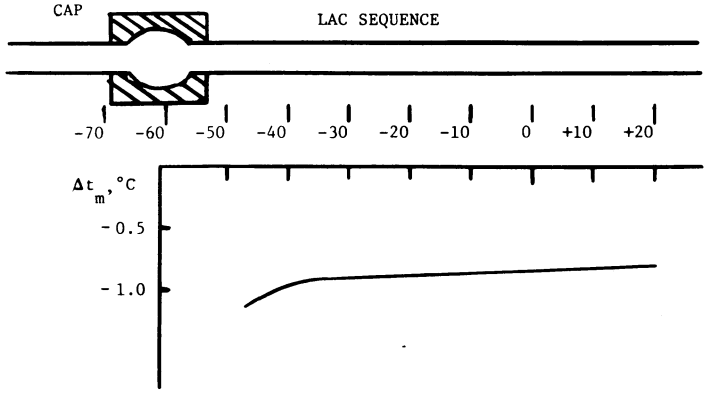


Figure 4: The change in the melting temperatures, Δt_m, of base pairs in the lactose operoninduced by a melting protein at the CAP site. See text for further details.

site on the transition midpoint of base pairs in the lac sequence. With no protein, all base pairs in the lac region gave essentially the same t_m, 89.3°C. In the presence of a melting protein the thermal instability adjacent to the melted region decreases slowly with distance. The t_m of base pair -47 is lowered by 1.07°C whereas for base pair +38 it is lowered by 0.77°C. The base pair opening rate of the whole region surrounding the melting protein appears to increase. This is also indicated by the increase in the transition width of the

base pairs melting curves (Figure 3). Estimates of the change in the average free energy of base pair formation, $d(\Delta G)$, at 35°C, were obtained from ΔT_m values. When no protein is present the entropy change, ΔS, was taken as -24.8 eu./mole base pair and ΔH was calculated at the measured T_m where $\Delta G = 0$. Assuming ΔS is constant, the new enthalpy change is calculated from $\Delta H^* = T_m^* \Delta S$ (T_m^* is the base pair transition midpoint with protein present). The changes in free energy, $d(\Delta G)$, at 35°C are estimated from $d(\Delta G) = \Delta H^*(1-T/T_m^*) - \Delta H(1-T/T_m)$. Table I lists values of $d(\Delta G)$ for several base pairs. The positive change indicates a decreased stability of the duplex state. The values are quite small, however over forty base pairs, -20 to +20, the total change in the free energy is about 810 calories. If one assigns the change in melting temperatures to equivalent changes in ΔH and ΔS one gets results very similar to Table I.

The influence of a melting protein centered at the mRNA initiation site of the lactose operon was also examined. This simulates the effect RNA polymerase has on the stability of adjacent pairs when it is poised in the "open complex." The ten base pairs from -4 to +5 were taken to be melted by RNA polymerase. The binding constant employed for the RNA polymerase-DNA interaction was $10^{12} M^{-1}$. The total RNA polymerase concentration was $5 \times 10^{-6} M$, and the non-specific binding parameters were the same as for the CAP protein. Table II shows how the "open complex" effects the stability of base pairs surrounding the transcription initiation site.

Base pairs -23 and +24 are both located 18 base pairs from the edges of the melted out region. Yet there is greater than a factor of two difference between their Δt_m. The t_m for base pair -23 is lowered by 1.04°C. The t_m of base pair +22 is lowered by 2.46°C. Without protein both base pairs have the same t_m, 89.28°C. This result shows the influence of base pair distribution

TABLE I: The effect of a melting protein at the CAP site on the stability of base pairs in the lactose operon. $d(\Delta G)$ is the change in the base pair free energy calculated at 35°C.

Base Pair	t_m, °C (no protein)	Δt_m, °C (+ protein)	$d(\Delta G)$ (cal/mole)
-47	89.28	-0.92	22.8
0	89.28	-0.86	21.3
+38	89.31	-0.77	19.1

in the region between the melting protein and the base pair being examined. There are 8/18 A·T pairs between sites -5 and -23 and 11/18 A·T pairs between sites +6 and +24.

2. The influence of base pair distribution on transmitted instability

The influence of base pair distribution in the region surrounding the protein site was examined further. The previous result shows that a small composition difference can generate a significant change in transmitted instability. The sequence of nucleotides appears to be a dominant part of this effect. When the 11/18 A·T pairs between sites +6 to +24 are rearranged such that there are no more than three consecutive A·T pairs (the new sequence is 00010100101010010 10 with 0 a A·T, 1 a G·C) the t_m decrease at base pair +24 is 1.14°C. There is no change in the t_m of base pair +24 when these two sequences are compared in the absence of a melted region.

Figure 5 demonstrates the large influence of the region intervening between the protein site and the base pairs being examined. A melting protein at the CAP site (-65 to -56) was examined with the fifteen base pairs -53 to -38 altered from the lac sequence. When this sequence was made 100% A·T the transmitted

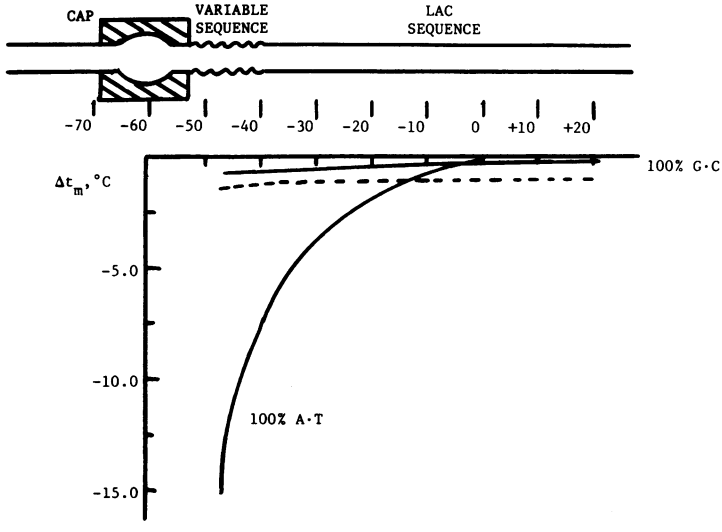


Figure 5: The change in t_m of base pairs in the lactose operon induced by a melting protein at the CAP site. The dashed line is the result obtained with the 15 base pair variable sequence the lactose operon sequence (Fig. 4). The 100% A·T curve corresponds to replacing the variable sequence with all A·T's. The 100% G·C curve is the result obtained with the variable sequence all G·C.

instability to the nearest 40 base pairs was greatly enhanced. Base pair -47 had a t_m of 70.9°C in the presence of melting protein. The t_m of base pair -37 was lowered by 6.8°C. Replacing this region with all G·C pairs reduced the t_m change to 0.52°C (90.72°C to 90.2°C) for the site -47. The Δt_m damps out faster when the variable sequence is all A·T pairs. At base pair 0, Δt_m is 0.31°C when the variable region is 100% G·C, and 0.1°C when this region is 100% A·T. Figure 5 shows that when an A·T rich region intervenes between a highly unstable site and a base pair, the transmitted instability is very large. When a G·C rich region intervenes, the transmitted change is small. These results can be physically understood as follows. An A·T rich section will tend to be thermally unstable. It stays duplexed because of stabilizing base pairs surrounding it. Placing a melting protein adjacent to it, destabilizes the base pairs keeping the A·T section duplexed. Thus the A·T section can open. The energy cost of expanding this loop a short distance is not too large. A G·C rich section will keep nearby base pairs duplexed. The energy cost of expanding a protein induced instability through the G·C section is high.

Figure 6 shows the influence of the base pair composition of the protein site on transmitted stability changes. In this case the ten base pairs centered

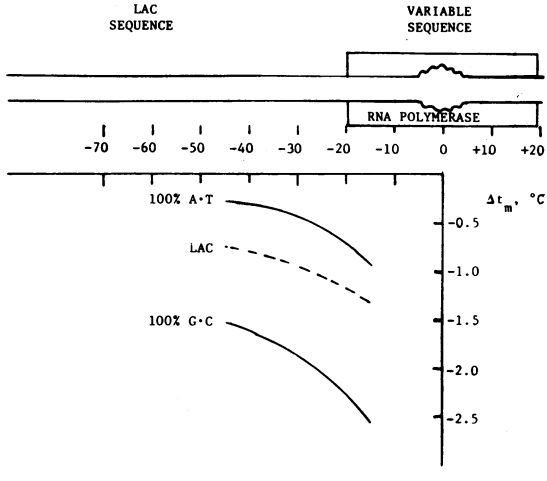


Figure 6: The effect of nucleotide composition of the melted site on transmitted instability. The Δt_m of base pairs in the lactose operon region induced by a melting protein opening base pairs -4 to +5. The region covered by RNA polymerase is also shown. The variable sequence is the 10 base pair melted site. The dashed line is the result obtained with this site the lactose operon sequence. 100% A·T and 100% G·C curves correspond to replacing this site with all A·T or all G·C pairs respectively.

at transcription initiation, the RNA polymerase site, was examined. This site, -4 to +5, was replaced by either 10 G·C pairs or 10 A·T pairs. The dashed line shows the transmitted instability when the lac sequence is used for the site (same as Table II). When the site contains only A·T pairs the transmitted instability decreases. For base pair -23 the t_m goes from 88.8°C to 88.27°C, $\Delta t_m = -0.53$°C. The lac sequence gives a Δt_m of -1.04°C for base pair -23. When the protein site contains only G·C pairs the transmitted instability increases. The t_m for base pair -23 now changes from 90.3°C to 88.27°C or -2.03°C. The influence of nucleotide composition in the melted site is opposite that of the intervening region. Whereas G·C pairs in the intervening region decrease transmission of instability, a G·C rich melted site increases the effect. This can be understood by arguments similar to those given before. A G·C rich site will tend to clamp itself and the surrounding region in the duplex state. Destabilizing this site makes it energetically cheaper to open up adjacent regions. An A·T rich region is already relatively unstable. Destabilizing such a site does not have very much effect on the surrounding region.

3. Stabilizing protein effect

The effect of a repressor protein on the surrounding DNA regions was also examined. The protein was considered to bind to the seventeen base pairs +3 to +19. The site specific binding constant was taken to be $K_1 = 10^{13} M^{-1}$ with a protein concentration of 2×10^{-8}M. Non-specific binding was $K_N = 10^6 M^{-1}$ and the concentration of non-specific sites, $S = 10^{-4}$M. Table III shows the influence of a stabilizing repressor on four distant base pairs.

Base pair -7 is stabilized by 1.1°C, b.p. -17 by 1.06°C and b.p. -47 by 0.56°C. These three base pairs are in the nontranscribed direction. Base pair +39 in the

TABLE II: Effect of melting protein at mRNA initiation site on surrounding DNA regions. Melted region is from -4 to +5.

Base Pair	t_m, °C (no protein)	Δt_m, °C (+ protein)
-13	89.28	-1.30
-23	89.28	-1.04
-43	89.28	-0.80
+24	89.28	-2.46
+38	89.31	-1.30

TABLE III:

Effect of stabilizing protein at repressor site on surrounding DNA region. Stabilized region is from base pairs +3 to +19.

Base Pair	t_m, °C (no protein)	Δt_m, °C (+ protein)
-7	89.28	+1.10
-17	89.28	+1.06
-47	89.28	+0.56
+39	89.31	+1.28

transcribed direction is stabilized by 1.28°C. A comparision between sites -7 and +39 again shows an effect of local base pair distribution on tele-stability. Although both base pairs are equidistant from the protein site, their stability depends on the intervening base pair distribution. The influence of base pair composition of the region surrounding the stabilizing protein was determined. The fifteen base pairs from -12 to +2 were made either all G·C or all A·T pairs. The melting temperature of base pair -17 was evaluated for these two sequences in the presence and absence of the stabilizing protein. When the intervening region was 100% A·T pairs, the t_m of base pair -17 without protein was 88.5°C, with protein t_m=89.1°C, or Δt_m = +0.6°C. When this region was 100% G·C pairs, the t_m of base pair -17 was 91.52°C without protein, and 99.76°C in the presence of protein or Δt_m=+8.24°C. Thus a G·C rich intervening region increases the transmission of stability for a stabilizing protein. These results are complimentary to the effect of nucleotide composition on telestability found for the melting proteins.

DISCUSSION

The results indicate that melting proteins decrease the stability of duplex regions 20-100 base pairs away from the protein-DNA complex in the helix-coil transition region. Stabilizing proteins transmit stability over similar distances. Recent results by Lukashin *et al.*[29] indicate that the correlation length for regional cooperativity decreases as a function of temperature. Their results are consistent with the small transmitted changes in the base pair free energies evaluated at 35°C. However, it must be kept in mind that the values are first order estimates. Interactions extending beyond neighboring base pairs, and the uncertainties of the parameter values may alter these estimates.

The base pair composition of the region surrounding the melted or stabilized site has a large influence on telestability. This is demonstrated by the results for melting proteins at the CAP site and the mRNA initiation site. When G·C pairs intervene between the melting protein and the base pairs examined, the transmitted instability is small. When A·T pairs intervene the transmitted instability is greatly enhanced.

The influence of nucleotide sequence adjacent to the protein site is demonstrated by the results obtained for a melting protein at the mRNA initiation site (-4 to +5). By slightly changing the sequence of base pairs in the region between base pairs +5 to +24, the Δt_m for base pair +24 was altered from -2.46°C to -1.14°C. The change in sequence has no effect on the t_m of the region's base pairs in the absence of a melting protein. Only the transmitted change in stability in the presence of a protein is altered. This change in Δt_m with sequence shows that base pair composition alone does not determine the melting behavior of base pairs near site specific binding proteins. These effects of nucleotide composition and sequence are also observed for the regions surrounding a stabilizing protein.

The base composition of a protein site also effects the transmission of instability or stability. This is shown in Fig. 6 for a melting protein at the mRNA initiation site. Altering the base pair composition of the site from the lac sequence showed that the transmission of instability is largest when a melted site is G·C rich and smallest when it is A·T rich. This behavior is opposite that for the region intervening between a melting protein site and a distant base pair.

Summarizing the above results, one can make the following conclusion. The transmitted instability from a melting protein is largest when the protein site is G·C rich and the intervening region is A·T rich. Conversely the transmitted change is smallest when the melted region is A·T rich and the intervening region is G·C rich. This observation may be relevant to the functional role of A·T/G·C blocks in DNA. Adjacent G·C rich and A·T rich regions have been found in DNA sequences; near the replication site of Φ X 174 virus DNA,[26] and in the promoter regions of the Escherichia coli tyrosine transfer RNA gene[27] and the lactose operon.[3] Also the 3' ends of several mRNAs terminated *in vitro* show a string of six to eight uracils preceded by a G·C rich sequence.[28] A discussion of the possible role of telestability in RNA polymerase termination will be presented elsewhere (manuscript in preparation).

One of the questions this study set out to examine was whether the CAP-cAMP complex, and positive effectors in general, can enhance RNA polymerase binding by transmitting thermal instability through DNA. The calculations show

that if CAP protein produces a melted region where it binds, it will decrease the energy required to open up the DNA region 40-80 base pairs away by about 800 calories. This region, -20 to +20 has been established as an RNA polymerase site by enzyme protection experiments[23] and electron microscopy[24]. 800 calories is about one third the activation energy required to explain the 50 fold increase in RNA polymerase activity due to CAP stimulation in vivo.[25] It may be argued that summing the free energy change over the entire 40 base pair region is unjustified, since not all base pairs will be interacting with polymerase. However, the possibility of intermediate states involving the entire region makes this qualm less certain. If the CAP stimulation is due to telestability, it will probably function by providing a low energy pathway between DNA conformations similar in equilibrium energy but separated by a large activation energy barrier.

ACKNOWLEDGEMENT

Support of this work by a grant (PCM76-04565) from the National Science Foundation is gratefully acknowledged.

REFERENCES

1. Von Hippel, P.H. (1969) J. Cell Physiol. 74, 235-241.
2. Burd, J.F., Wartell, R.M. Dodgson, J.B. and Wells, R.D. (1975) J. Biol. Chem. 250, 5109-5113.
3. Dickson, R.C., Abelson, J., Barnes, W.M. and Reznikoff, W.S.(1975) Science 187, 27-35.
4. Burd, J.F.,Larson, J.E.,and Wells, R.D.(1975) J. Biol.Chem. 250,6002-6007.
5. Wartell, R.M. and Burd, J.F. (1976) Biopolymers 15, 1461-1479.
6. Bloomfield, V., Crothers D.M. and Tinoco, Jr. I. (1974) Physical Chemistry of Nucleic Acids Harper and Row, New York.
7. Borer, P.N. Dengler, B. Tinoco, Jr., I. and Uhlenbeck, O.C. (1974) J. Mol. Biol. 86, 843-853.
8. Wartell, R.M. and Montroll, E.W. (1972) Adv. Chem. Phys. 22, 130-201.
9. Crothers, D.M. (1969) Accts. Chem. Res. 2, 225-231.
10. Poland, D. and Scheraga, H.A.(1970) Theory of the Helix Coil Transition in Biopolymers, Academic Press, New York.
11. Poland D. (1974) Biopolymers 13 1859-1871.
12. Lukashin A.V., Vologodskii A.V. and Frank-Kamenetski M.D. B(1976) Biopolymers 15, 1841-1844.
13. Scheffler, I.E. and Sturtevant, J. (1969) J. Mol. Biol. 42 577-581.
14. Oliver, A.L., Wartell R.M. and Ratliff (1977) Biopolymers in press.
15. Wartell, R.M., Oliver A.L. and Howell D.K. manuscript in preparation.
16. Gralla J., and Crothers, D.M. (1973) J. Mol. Biol. 73, 497-511.
17. Crothers, D.M. (1971) Biopolymers 10, 2147-2160.
18. Schellman J. (1975) Biopolymers 14, 999-1018.
19. McGhee J.D. (1976) Biopolymers 15, 1345-1375.
20. Lin, S. and Riggs A.D. (1972) J. Mol. Biol. 72, 671-690.
21. Wu, F.Y. Nath, K. and Wu C.W. (1974) Biochemistry 13, 2567-2572.
22. Riggs, A.D.,Reiness G.and Zubay G.(1971)Proc.Nat.Acad. Sci. 68, 1222-1225.
23. Gilbert, W. and Maxam A. (1973) Proc. Nat. Acad. Sci 70, 3581-3584.
24. Hirsch, J. and Schleif R. (1976) J. Mol. Biol. 108, 471-490.

25. Majors J. (1975) Nature 256, 672-673.
26. Sanger, F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R. Fiddes J.C., Hutchinson C.V., Slocombe P.M. and Smith, M. (1977) Nature 265, 687-694.
27. Sekiya, T. Gait, M. Noris, K. Ramamoorthy, B. and Khorana, H.G. (1976) J. Biol. Chem. 251, 4481-4489.
28. Gilbert. W. (1976) in RNA Polymerase, pp. 193-206 (ed. R. Losick and M. Chamberlin) Cold Spg. Hrbr. Lab. New York.
29. Lukashin A.V., Vologodskii, A.V., Frank-Kamenetskii M.D., and Lyubschenko, Y.L. (1976) J. Mol. Biol. 108, 665-682.