ELSEVIER

Biophysical Chemistry 125 (2007) 497–507

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Enthalpy distribution functions for protein-DNA complexes: Example of the binding of AT-hooks to target DNA

Douglas Poland *

Department of Chemistry, The Johns Hopkins University, Baltimore, MD 21218, USA

Received 9 October 2006; received in revised form 29 October 2006; accepted 30 October 2006
Available online 20 November 2006

Abstract

In this article we use the published heat capacity data of Dragan et al. [A.I. Dragan, et al., The energetics of specific binding of AT-hooks from HMGA1 to target DNA, J. Mol. Biol. 327 (2003) 393–411] on the association of proteins with DNA duplexes to construct enthalpy probability distributions for the protein/DNA complexes formed in these systems. We first analyze the multistep equilibrium that determines the species concentrations in this system to determine whether or not the DNA-peptide complex goes cleanly to DNA single-strands and peptide. Using the heat capacity data for this case we employ the maximum-entropy method to construct enthalpy probability distribution functions for the species involved in this equilibrium. We find that the distribution functions for this system clearly show bimodal behavior indicating a two-state transition from complex to non-complex form.

Keywords: Enthalpy distribution; Protein-DNA complex; Thermal denaturation; Maximum-entropy method; Bimodal distributions; Two-state behavior

1. Introduction

In the present paper we use the heat capacity data of Dragan et al. [1] on the association of proteins and DNA to give complexes in order to determine enthalpy probability distributions for these systems. In a previous publication [2] we used the heat capacity data of these authors to determine enthalpy distribution functions for short DNA molecules. The calculation of such enthalpy distributions [3–8] utilizes the temperature dependence of the heat capacity to calculate moments of the enthalpy distribution, which in turn are used to construct the enthalpy probability distribution as a function of temperature. A review of the method is available [9].

The systems studied by Dragan et al. [1] are of the form

$$\text{peptide} + \text{DNA} \leftrightarrow \text{complex} \qquad (K_a) \tag{1}$$

where K_a is the equilibrium constant for the association reaction. Specifically, they studied several small DNA-binding protein motifs that bind selectively to AT-rich DNA sequences by penetrating deep into the minor grove without significant distortion of the DNA duplex. Because of their affinity for AT-rich DNA-binding sites, these protein motifs are referred to as AT-hooks. We will examine the data for one of the peptides used by Dragan et al., namely the DBD2 peptide. This is a 17-residue peptide the structure of which is given below:

$$N- \text{ acetyl} - \underset{\text{(DBD2 peptide)}}{\text{PKRPRGRPKG} - \text{SKNKGAA}} - \text{NH}_2 \tag{2}$$

We will refer to this molecule simply as “the peptide”. The target DNA used was composed of 12 base pairs with the following sequence:

$$\begin{array}{l} 5'- \text{ G G G A A A T T C C G C} - 3' \\ 3'- \text{ C C C T T T A A G G C G} - 5' \end{array} \qquad \underset{\text{(target DNA)}}{} \tag{3}$$

The DNA duplex given above corresponds to one of the naturally occurring binding sites of the DBD2 hook motif. The DBD2 sequence is one of three different hook motifs.

* Tel.: +1 410 516 7441; fax: +1 410 516 8420.
E-mail address: poland@jhu.edu.

0301-4622/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2006.10.012

The overall reaction scheme for this system can be written as a step-wise dissociation process as follows:

$$\text{(reaction-1)} \qquad C \overset{x}{\longleftrightarrow} D + P \qquad (K_1 = 1/K_a) \tag{4}$$

$$\text{(reaction-2)} \qquad D \overset{y}{\longleftrightarrow} S_a + S_b \qquad (K_2) \tag{5}$$

In the above reactions C stands for the DNA-peptide complex, D stands for the DNA duplex, P stands for the peptide while S_a and S_b represent the two complementary single-strands of DNA formed on dissociation of the DNA duplex. We note in Eq. (4) that this is the reverse of the association reaction given in Eq. (1) and that $K_1 = 1/K_a$. We can add these two reactions together to give the overall reaction for the dissociation of the complex into its component parts as follows

$$\text{(reaction-3)} \qquad C \longleftrightarrow S_a + S_b + P \qquad (K_3 = K_1 K_2) \tag{6}$$

The equilibrium constants for the reactions are K_1, K_2, and K_3 respectively. As we have indicated above, only two of the above three reactions are independent. The quantities x and y indicated in Eqs. (4) and (5) are the respective progress variables for the two reactions. We note that in the reaction scheme given above we have assumed that the peptide does not bind to the single-strands of DNA.

In Fig. 1 we show heat capacity data of Dragan et al. [1] for the complex formed from the peptide and DNA indicated respectively in Eqs. (2) and (3) for the case of $r = 3.5$ (the parameter r is the ratio of the initial peptide concentration to the initial DNA concentration). The heat capacities of the DNA without any peptide present and the peptide without any DNA present are also shown. In the present paper we will use the temperature dependence of the heat capacity for the complex shown in Fig. 1 to construct the enthalpy probability distribution for this species. To do this we first need to determine which of the species appearing in Eqs. (4)–(6) are dominant as a function of temperature and whether reaction-1 or reaction-3 dominates. We can do this by using the data of Dragan et al. [1]

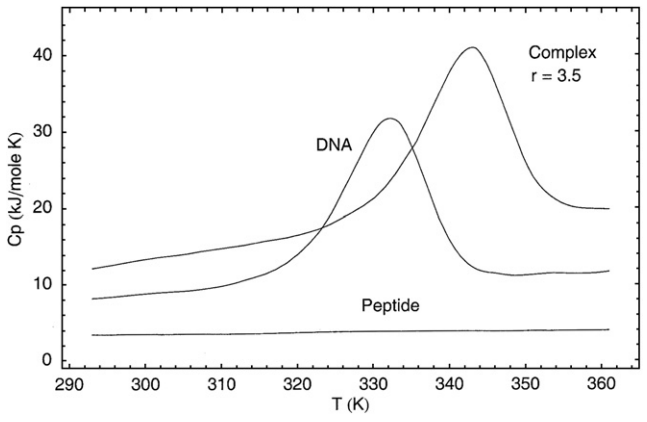

Fig. 1. The heat capacity as given by Dragan et al. [1] for the molecules described in (2)–(3). The curves marked "DNA" and "Peptide" respectively give the heat capacity for these isolated systems (with concentration 180 μm for each). The curve marked "complex" gives the heat capacity for a mixture of the two species (with DNA concentration of 180 μm and peptide concentration 3.5 times this value).

to estimate K_1, K_2, and K_3 in the above reaction scheme. We begin by specifying the coupled equilibrium scheme in Eqs. (4)–(6) in more detail.

2. Cooperative equilibria

To describe the reactions given in Eqs. (4)–(6) more fully we first need to specify the concentrations of the species involved. The most important parameter is the total DNA concentration in any form, which we denote as c_o. For the present system Dragan et al. [1] give

$$c_o = 180\ \mu M \tag{7}$$

We let p_o represent the total concentration of the peptide in any form (bound or free). It is given by the following relation

$$p_o = r\ c_o \geq c_o \tag{8}$$

where Dragan et al. give data for $r = 1$, 1.5, 2, 3.5 and 5. In terms of the initial concentrations given in Eqs. (7) and (8) and the progress variables defined in Eqs. (4) and (5) we have the following relations for the concentrations of all of the species involved in this system (the square brackets indicate concentration, conventionally in μM):

$$[C] = c_o - x, \quad [D] = x - y, \quad [P] = (r-1)c_o + x, \quad [S_a] = [S_b] = y = [S] \tag{9}$$

Since the maximum concentration of peptide that can bind is c_o, the amount $(r-1)\ c_o$ remains in solution as a background concentration.

For the reactions under consideration we have two conservation relations:

$$[C] + [P] = (c_o - x) + (r-1)c_o + x = p_o \quad \text{(conservation of peptide)} \tag{10}$$

$$[C] + [D] + [S] = (c_o - x) + (x - y) + y = c_o \quad \text{(conservation of DNA)} \tag{11}$$

The equilibrium constant expressions for the three reactions in Eqs. (4)–(6) are then given in terms of the variables expressed in Eq. (9) as follows

$$K_1 = \frac{[D][P]}{[C]} = \frac{((r-1)c_o + x)(x-y)}{(c_o - x)}, \quad K_2 = \frac{[S_a][S_b]}{[D]} = \frac{y^2}{(x-y)}$$

$$K_3 = \frac{[S_a][S_b][P]}{[C]} = \frac{((r-1)c_o + x)y^2}{(c_o - x)} \tag{12}$$

We note that K_1 and K_2 have the units of μM while K_3 has the units of $(\mu M)^2$.

We next introduce the following dimensionless fractional quantities

$$[P]/c_o = (r-1) + f, \quad [C]/c_o = 1 - f \tag{13}$$

$$[S_a]/c_o = [S_b]/c_o = g, \quad [D]/c_o = f - g$$

where f and g are the scaled progress variables

$$f = x/c_o \quad \text{and} \quad g = y/c_o \tag{14}$$

In terms of the fractional quantities given in Eq. (13), the conservation relations of Eqs. (10) and (11) become

$$(r-1) + f + (1-f) = r \quad \text{(conservation of peptide)} \tag{15}$$

$$(1-f) + (f-g) + g = 1 \quad \text{(conservation of DNA)} \tag{16}$$

From Eq. (13) we note that

$$[D]/c_o = f-g \tag{17}$$

In order to have [D] be a positive quantity the following relation must hold

$$f \geq g \tag{18}$$

Thus we have the following set of inequalities for the two progress variables

$$1 \geq f \geq g \geq 0 \tag{19}$$

We now introduce the following dimensionless scaled equilibrium constants

$$k_1 = K_1/c_o,\ k_2 = K_2/c_o,\ k_3 = K_3/c_o^2 \tag{20}$$

Using the scaled variables given in Eqs. (13) and (20), the equilibrium constantexpressions of Eq. (12) then become

$$k_1 = \frac{(r-1+f)(f-g)}{(1-f)},\ k_2 = \frac{g^2}{(f-g)},\ k_3 = \frac{g^2(r-1+f)}{1-f} \tag{21}$$

where we note that these expressions are all given in terms of dimensionless variables.

In Eq. (20), the lower-case k's are related to the upper-case K's by division by the quantity c_o, the total DNA concentration. If we want to treat a different DNA concentration, say c_o', then we have the following scaling

$$k_1' = \left(\frac{c_o}{c_o'}\right)k_1,\ k_2' = \left(\frac{c_o}{c_o'}\right)k_2,\ k_3' = \left(\frac{c_o}{c_o'}\right)^2 k_3 \tag{22}$$

3. Equilibrium constants and species plots

If one knows the standard enthalpy difference between products and reactants for a general reaction expressed as a series in the absolute temperature

$$\Delta H^o = a + bT + cT^2 \tag{23}$$

then one can obtain the equilibrium constant for the reaction as a function of temperature, $k(T)$, from the van't Hoff equation

$$\frac{\partial \ln k(T)}{\partial (1/T)} = -\Delta H^o/R. \tag{24}$$

Using the form for ΔH^o given in Eq. (23) one can integrate Eq. (24) from a reference temperature T_r to a general T giving

$$\ln k(T) = \ln k(T_r) - \frac{1}{R}\left[a\left(\frac{1}{T}-\frac{1}{T_r}\right) + b\ \ln\left(\frac{T_r}{T}\right) + c(T_r-T)\right] \tag{25}$$

where $k(T_r)$ is the value of the equilibrium constant at the reference temperature T_r. Thus knowledge of the series for ΔH^o given in Eq. (23) and the value of the equilibrium constant at one temperature is enough to give a good estimate of $k(T)$.

Dragan et al. [1] explicitly measured the ΔH of the association reaction given in our Eq. (1); their results are contained in their Eq. (1). We note again that this reaction is the reverse of our reaction-1 given in Eq. (4) with the following relation between the equilibrium constants for the two reactions

$$K_1 = 1/K_a \tag{26}$$

Thus to obtain the ΔH for our reaction-1 we reverse the sign of their ΔH (and switch units from Centigrade to Kelvin). The result is (giving ΔH in kJ/mol)

$$\Delta H_1^o = -177 + 0.6T \tag{27}$$

In addition they give numerical values of K_a at 10, 20, and 30 C in their Table (1). In a private communication Prof. Privalov has noted that the factor of 10^{-3} that multiples K in their Table (1) should not be there. They note that their version of Eq. (27) above gives the most accurate representation of the temperature dependence of K_a and hence we will use that relation. In addition we need the value of K_a at one temperature and we take the value given in their Table (1) for 30 C which is $K_a = 10^{5.220}\ M^{-1}$. Using Eqs. (20) and (26) and the value of c_o given in Eq. (7) expressed as $c_o = 10^{-3.745}$ M we obtain the dimensionless quantity

$$k_1(T_r) = 10^{-1.475} \tag{28}$$

where

$$T_r = 30 + 273.15 = 303.15\ K \tag{29}$$

Using Eqs. (27), (28) and (29) in Eq. (25) gives a good estimate of $k_1(T)$.

In a previous publication [2] we have determined the temperature dependence of the equilibrium constant for reaction-2 given in Eq. (5). We find that the standard enthalpy difference for this reaction is given by the following relation

$$\Delta H_2^o = -2297 + 16.23\ T - 0.0255\ T^2 \tag{30}$$

In addition, for this reaction we have

$$T_r = 332.2\ K \tag{31}$$

$$k_2(T_r) = 1/2$$

Using the data of Eqs. (30) and (31) in Eq. (25) gives $k_2(T)$.

Given the temperature dependence of the equilibrium constants for reaction-1 and reaction-2 of Eqs. (4) and (5), we then use the relation given in Eq. (6) to obtain the equilibrium

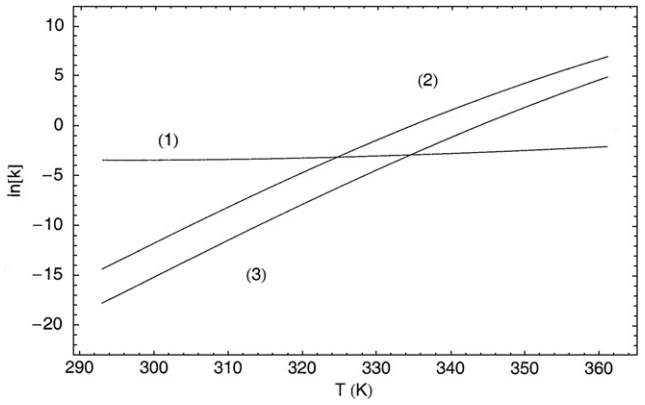

Fig. 2. The temperature dependence of the equilibrium constants for the three reactions described in Eqs. (4)–(6) in the dimensionless form given in Eq. (20) with c_o=180 μm. The constants were constructed using the data contained in Eqs. (23)–(32).

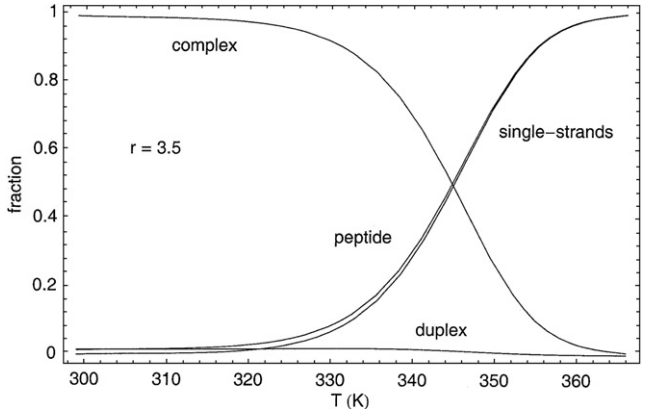

Fig. 3. The concentrations of the species involved in the complex-formation reactions given in Eqs. (4)–(6) expressed as the fractional quantities defined in Eq. (13). The concentrations were calculated using Eqs. (33) and (34). The calculations are for the case where c_o=180 μm and r=3.5. The peptide curve gives the peptide fraction minus (r−1).

constant for reaction-3. Using the scaled forms of Eq. (20) we have

$$k_3(T) = k_1(T)k_2(T) \tag{32}$$

The variation of k_1, k_2, and k_3 with temperature, expressed as the natural logarithms of the constants, is shown in Fig. 2. One notes that k_1 is almost independent of temperature and that k_2 and k_3 parallel one another as a function of temperature. In comparing k_2 and k_3 it is important to recall that reaction-2 is bimolecular while reaction-3 is trimolecular.

Given the temperature dependence of the k's we can then use Eq. (21) to give the species concentrations as a function of temperature. We can use any two of the three equations given (as indicated in Eq. (6), only two of the reactions are independent) and we choose to use reaction-1 and reaction-2. We first use the equilibrium constant expression for reaction-2 to solve for f as a function of g giving

$$f = g + g^2/k_2 \tag{33}$$

We then substitute this into the equilibrium constant expression for reaction-1 giving the following quartic equation in the single variable g

$$g^4 + k_2 g^3 + (k_1 + r - 1)\, k_2 g^2 \; + \; k_1 k_2^2\, g - k_1 k_2^2 = 0 \tag{34}$$

The solution of this equation in the physically meaningful range $0<g<1$ gives the unique solution of Eq. (34) for this system. The corresponding value of f is then given by Eq. (33). This set of values, (f, g), are then used in Eq. (13) to give the equilibrium concentrations of all of the species involved in this system as a function of temperature.

This variation of the species concentrations is shown in Fig. 3 using the value of c_o given in Eq. (7) and the value of r (as defined in Eq. (8)) taken as r=3.5. The fractional concentration of peptide plotted in Fig. 3 is the free peptide minus the excess background concentration. From Eq. (13) this is

$$[P']/c_o = f \tag{35}$$

The striking result of this calculation is that reaction-3 as given in Eq. (6) is the dominant process going on. That is, the peptide-DNA complex dissociates directly to the single-strands of DNA and the released peptide. Free DNA in the duplex form never makes up any significant part of the process. The reason for this is simply that the mass action effect of a large peptide concentration stabilizes the peptide-DNA complex to a temperature that is higher than the melting temperature of the DNA alone and hence the complex dissociates directly to the peptide and the DNA single-strands. For comparison we show in Fig. 4 the calculated species population profiles for the case r=1.0 (with all the rest of the parameters the same). In this case the amount of the complex remains roughly constant at about 80% (with about 20% DNA duplex) over a broad temperature range until it finally dissociates to single-strands at higher temperatures. Thus the amount of peptide is very important in determining the species population profiles as a function of temperature. Based on the results shown in Fig. 3 for the case where r=3.5 (indicating that reaction-3 is the dominant

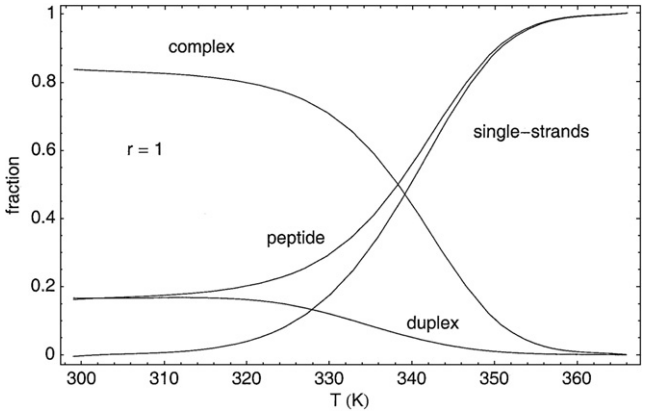

Fig. 4. The concentrations of the species involved in the complex-formation reactions given in Eqs. (4)–(6) expressed as the fractional quantities defined in Eq. (13). The concentrations were calculated using Eqs. (33) and (34). The calculations are for the case where c_o=180 μm and r=1.

process) we will interpret the enthalpy distributions that we obtain solely in terms of the amount of complex, single-strands of DNA and free peptide, with duplex DNA playing only a very minor role in the equilibrium.

4. Alternate calculation of k_3

Using the estimates of the three equilibrium constants given in Eqs. (23)–(32) we can calculate the mole fractions of all of the pertinent species as shown in Fig. 3. We now use the result shown there, namely that reaction-3 is the dominant one and that the main process going on is the dissociation of the DNA-peptide complex directly into the peptide and single-strands, to give an alternate calculation of the equilibrium constant k_3 for reaction-3. We begin by using the heat capacity data given in Fig. 1. Recall that the data given there is for the case where $r=3.5$ which means that there is a background concentration of free peptide equal to 2.5 c_o that never takes part in complex formation. We thus subtract out the peptide background heat capacity (which is 2.5 times the value of the peptide heat capacity given in Fig. 1) from the heat capacity of the complex. Since the heat capacity of the peptide is almost constant as a function of temperature, in reality this procedure simply shifts the heat capacity of the complex down by a constant amount.

The resulting heat capacity curve for the complex is shown by the solid curve in Fig. 5. Given $C_p(T)$ one can then calculate the corresponding enthalpy and entropy curves using the standard integrals

$$H(T)=\int_{T_o}^{T} C_p(T)\mathrm{d}T \quad \text{and} \quad S(T)=\int_{T_o}^{T}\frac{C_p(T)}{T}\mathrm{d}T \tag{36}$$

which give, respectively, the enthalpy and entropy of the system relative to the values at T_o (which is the lowest value of the temperature for which one has heat capacity data, 293 K in the present case). To keep the notation as simple as possible we refer to the quantities given in Eq. (36) as $H(T)$ and $S(T)$ rather than as

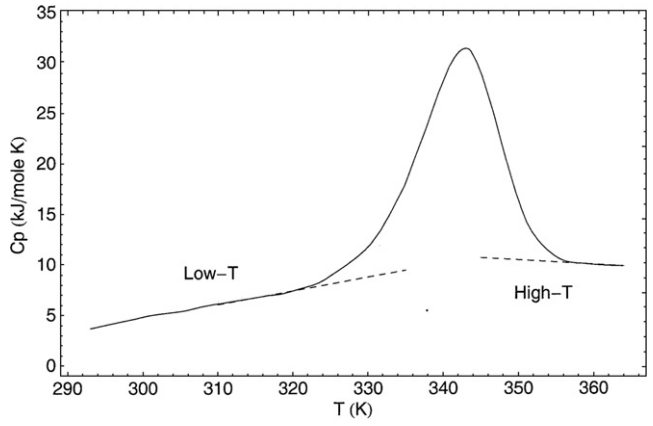

Fig. 5. The solid curve gives the heat capacity of the DNA-peptide system shown in Fig. 1 reduced by subtraction of the heat capacity for the excess peptide concentration. The dashed curves give the Low-T (complex) and High-T (DNA single-strands plus peptide) branches given by Eqs. (38)–(40).

$\Delta H(T)$ and $\Delta S(T)$. The Gibbs free energy is then given by combining $H(T)$ and $S(T)$ in the usual manner

$$G(T)=H(T)-T\ S(T) \tag{37}$$

which, like $H(T)$ and $S(T)$, is G relative to the value at T_o.

We note that the reduced $C_p(T)$ curve shown in Fig. 5 exhibits linear behavior with respect to the temperature at the low and high temperature ends of the $C_p(T)$ curve, on either side of the heat capacity maximum. We interpret these linear ranges as representing, respectively, the complex at low temperatures and the peptide plus single-strands at high temperatures, as indicated in Fig. 3. The maximum in the $C_p(T)$ curve represents the transition from the complex to the peptide plus single-strands that occurs at intermediate values of the temperature. We fit the low and high temperature ranges of the $C_p(T)$ curve to linear functions of the form

$$(\text{Low}-\text{T}) \qquad C_L(T)=a_L+b_L\ T \tag{38}$$

$$(\text{High}-\text{T}) \qquad C_H(T)=a_H+b_H\ T \tag{39}$$

where

$$\begin{aligned} &a_L=-35.98,\ a_H=26.16\\ &b_L=0.1355,\ b_H=-0.04497 \end{aligned} \tag{40}$$

The units for the constants given in Eq. (40) are such as to give C_p in kJ/(mol K). The linear branches given by Eqs. (38) and (39) are plotted (dashed lines) in Fig. 5.

One can then use the heat capacity branches given above to calculate the Low-T and High-T branches for the enthalpy as follows:

$$(\text{Low}-\text{T}) \quad H_L^o(T)=\int_{T_L}^{T} C_L(T)\ \mathrm{d}T \tag{41}$$
$$=H(T_L)+a_L(T-T_L)+\frac{1}{2}\ b_L(T^2-T_L^2)$$

$$(\text{High}-\text{T}) \quad H_H^o=\int_{T_H}^{T} C_H(T)\ \mathrm{d}T \tag{42}$$
$$=H(T_H)+a_H(T-T_H)+\frac{1}{2}\ b_H(T^2-T_H^2)$$

where

$$\begin{aligned} &T_L=293\ \text{K},\ T_H=356\ \text{K},\ H(T_L)=0,\\ &H(T_H)=769\ \text{kJ/mol} \end{aligned} \tag{43}$$

The corresponding entropy branches are obtained in a similar manner:

$$(\text{Low}-\text{T}) \qquad S_L^o=\int_{T_L}^{T}\frac{C_L(T)}{T}\mathrm{d}T \tag{44}$$
$$=S(T_L)+a_L\ln\left(\frac{T}{T_L}\right)+b_L(T-T_L)$$

$$(\text{High}-\text{T}) \qquad S_H^o=\int_{T_H}^{T}\frac{C_H(T)}{T}\mathrm{d}T \tag{45}$$
$$=S(T_H)+a_H\ln\left(\frac{T}{T_H}\right)+b_H(T-T_H)$$

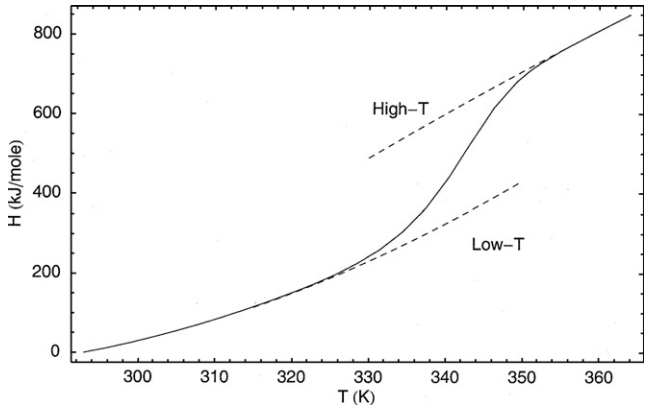

Fig. 6. The solid curve gives the relative enthalpy of the DNA-peptide system as obtained from the appropriate heat capacity integral of Eq. (36). The dashed curves give the Low-T (complex) and High-T (DNA single-strands plus peptide) branches given by Eqs. (41)–(43).

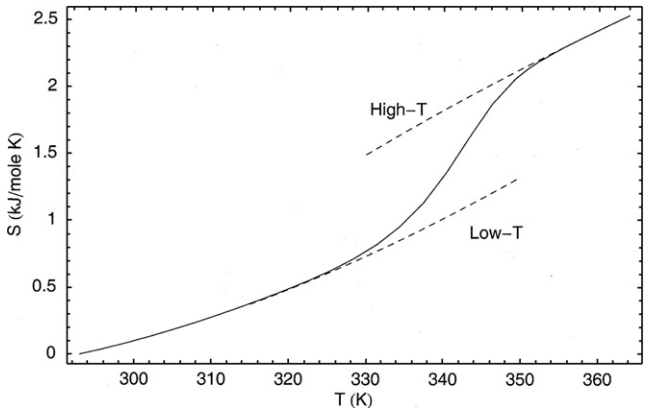

Fig. 7. The solid curve gives the entropy of the DNA-peptide system as obtained from the appropriate heat capacity integral of Eq. (36). The dashed curves give the Low-T (complex) and High-T (DNA single-strands plus peptide) branches given by Eqs. (44)–(46).

where

$$S_L = 0 \quad \text{and} \quad S_H = 2.308 \text{ kJ/(mol K)} \tag{46}$$

The Low-T and High-T branches for the enthalpy given by Eqs. (41) and (42) are plotted in Fig. 6 (dashed curves). The solid curve in Fig. 6 is the average enthalpy as given in Eq. (36). The Low-T enthalpy branch is for the complex while the High-T enthalpy branch is for the peptide plus single-strands of DNA; the average enthalpy is a mixture of the two. The corresponding Low-T and High-T branches for the entropy as given by Eqs. (44) and (45) are plotted in Fig. 7 (dashed curves) while the solid curve in Fig. 7 is the average entropy as given in Eq. (36). Again the Low-T entropy branch is for the complex while the High-T entropy branch is for the peptide plus single-strands of DNA.

Given the Low-T and High-T branches for the enthalpy and the entropy we can then construct the corresponding branches for the free energy. One has

$$\begin{array}{ll} (\text{Low}-\text{T}) & G_L^o(T) = H_L^o(T) - T\,S_L^o(T) \\ (\text{High}-\text{T}) & G_H^o(T) = H_H^o(T) - T\,S_H^o(T) \end{array} \tag{47}$$

These two branches are shown by the dashed curves in Fig. 8 while the solid curve gives the average $G(T)$ as given by Eq. (37). The net ΔG^o for reaction-3 of Eq. (6) is then given by

$$\Delta G_{HL}^o(T) = G_H^o(T) - G_L^o(T) \tag{48}$$

which is plotted in Fig. 9. This quantity then gives an alternate estimate of the equilibrium constant k_3 for reaction-3 of Eq. (6) in scaled, dimensionless units. This alternate form, which we designate with a prime, is given by the standard relation

$$k_3' = \exp[-\Delta G_{HL}^o(T)/RT] \tag{49}$$

Alternatively one can construct the standard enthalpy difference

$$\Delta H_{HL}^o(T) = H_H^o(T) - H_L^o(T) \tag{50}$$

and then use the van't Hoff equation of Eq. (25) with

$$T_r = 341.0 \text{ K}, \quad k_3'(T_r) = 1 \tag{51}$$

The T_r given in Eq. (51) is the value of the temperature where $\Delta G^o(T) = 0$. One obtains identical results using either Eq. (48) or Eq. (50) to calculate k_3'.

We can now compare the results obtained using k_3 with those obtained using k_3'. First we note from Fig. 3 that for $r = 3.5$ the DNA duplex never has any significant probability, that is $[D]/c_o \approx 0$. Thus from Eq. (17) we have the result that $f = g$ and hence we can write reaction-3 of Eq. (6) as

$$\begin{array}{ccccccc} C & \longleftrightarrow & S_a & + & S_b & + & P \\ 1-f & & f & & f & & (r-1)+f \end{array} \tag{52}$$

The equilibrium constant expression for this reaction is then

$$k_3 = \frac{f^2(r-1+f)}{1-f} \tag{53}$$

which we then can solve for f (with $r = 3.5$). An identical relation holds for k_3'.

In the upper graph in Fig. 10 we compare the fractions (f = fraction of non-complex, $(1-f)$ = fraction of complex)

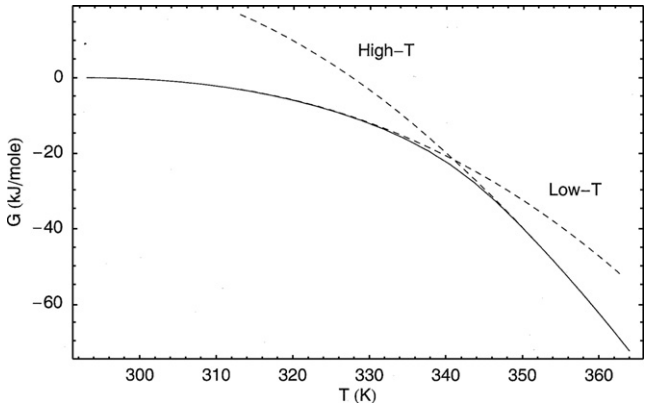

Fig. 8. The solid curve gives the Gibbs free energy of the DNA-peptide system obtained from Eq. (37) where $H(T)$ and $S(T)$ are calculated using the heat capacity integrals of Eq. (36). The dashed curves give the Low-T (complex) and High-T (DNA single-strands plus peptide) of the free energy of the DNA-peptide system as given by Eq. (47).

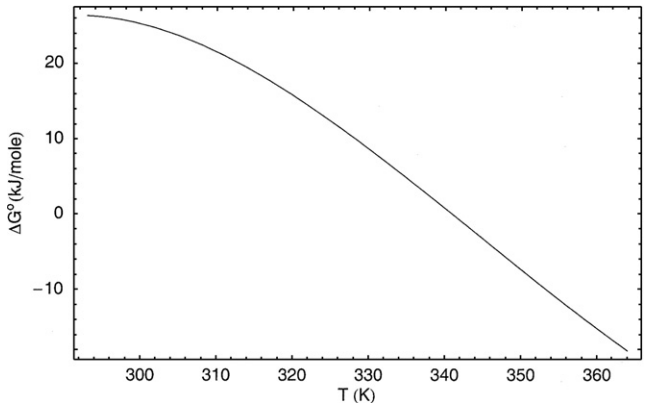

Fig. 9. The standard free energy change for reaction-3 of Eq. (6) as obtained using the free energies of Eq. (47) in Eq. (48).

calculated using both k_3 and k_3'. The solid curves show the results obtained by using k_3 (calculated from Eq. (32)) while the dashed curves give the results obtained by using k_3' (calculated from Eq. (49)); in both cases the fractions were calculated using Eq. (53). While the results obtained from the two equilibrium constants are not identical, they are clearly giving the same overall picture.

Another way to estimate the fractions of complex and non-complex is to use the enthalpy curves shown in Fig. 6 and interpolate the value of the quantity f. The average enthalpy, $H(T)$, is given in terms of the two enthalpy branches, $H_L^o(T)$ and $H_H^o(T)$, as follows

$$H(T) = (1-f)H_L^o(T) + f\ H_H^o(T) \tag{54}$$

Solving this relation for f one obtains

$$f = \frac{H(T)-H_L^o(T)}{H_H^o(T)-H_L^o(T)} \tag{55}$$

Alternatively one can construct an equivalent relation using the entropy curves shown in Fig. 7. The results obtained from using Eq. (55) are shown by the solid curves in the lower graph in Fig. 10. The dashed curves repeat the results obtained using k_3' in Eq. (53) as shown in the upper graph. In summary, Fig. 10 shows three different ways to calculate the fractions of complex and non-complex: using k_3 of Eq. (32) in Eq. (53), using k_3' of Eq. (49) in Eq. (53) and, finally, using Eq. (55). While there are moderate differences in the results obtained (mainly due to uncertainties in the equations for the heat capacity branches given in Eqs. (38)–(40)), all three approaches give the same general picture for the temperature variation of complex and non-complex.

We note that the expectation that the three approaches for calculating the species probabilities should give the same result is based on the following assumptions. First, that the general reaction scheme outlined in Eqs. (4)–(6) accurately represents the chemistry taking place. Second, that for the special case of $r=3.5$ the duplex concentration is essentially zero (as indicated in Fig. 3) and hence the chemistry reduces to the reaction shown in Eq. (52) with the equilibrium constant expression given by Eq. (53). And third, that the asymptotic branches for the heat capacity given in Eqs. (38)–(40) are accurate. Given the correct chemistry and the correct branches, one can then calculate the main thermodynamic functions (H, S and G) as given in Figs. 6–8. The comparison of the calculated thermodynamics with the experimental thermodynamics (obtained from the heat capacity as indicated in Eq. (36)) gives a check on the validity of these assumptions. We illustrate this approach by using k_3' in Eq. (53) (which requires knowledge of the correct chemistry and the branches of G) to calculate f which is then used in Eq. (54) (which requires knowledge of the branches of H) to finally give $H(T)$. The enthalpy calculated in this manner is shown by the dashed curve in Fig. 11 while the solid curve gives the experimental variation of $H(T)$ as given in Fig. 6. One sees that the calculated and empirical values of $H(T)$ are in good agreement, thus supporting our basic assumptions.

5. Enthalpy distribution functions

We now use knowledge of the temperature dependence of the heat capacity to construct enthalpy distribution functions for this system. We have described this method in detail in previous publications [2,3,5–8]; a review is available [9]. Here we very briefly sketch the method. The quantity that is available experimentally from integration of the heat capacity is the enthalpy, $H(T)$, as given by Eq. (36). In order to calculate the enthalpy distribution function we require moments of this

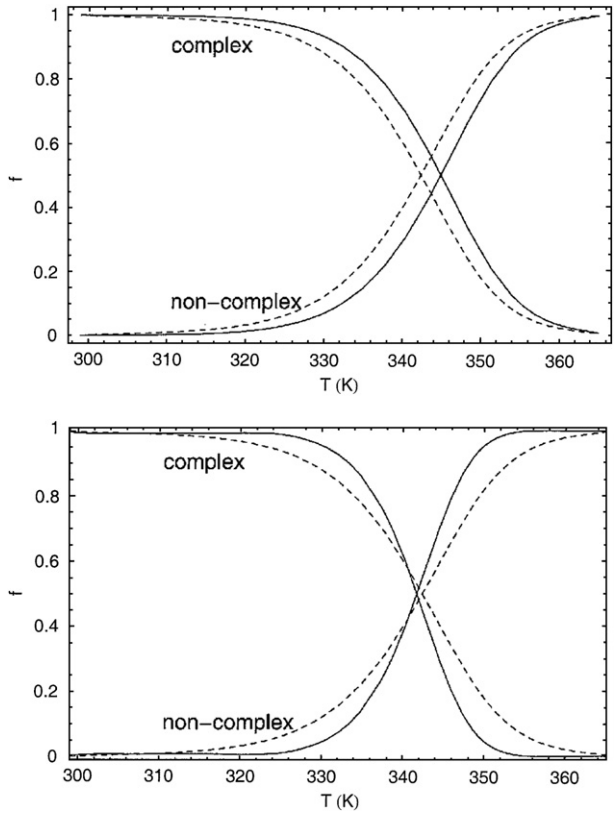

Fig. 10. The extent of reaction-3 as given by Eq. (52) in terms of the variable f (designated here as "fraction"). The dashed curves in both graphs are obtained from k_3' given by Eq. (49) and used in Eq. (53). The solid curves in the upper graph are obtained from k_3 given by Eq. (32) and used in Eq. (53) while the solid curves in the lower graph are obtained from the interpolation relation of Eq. (55).

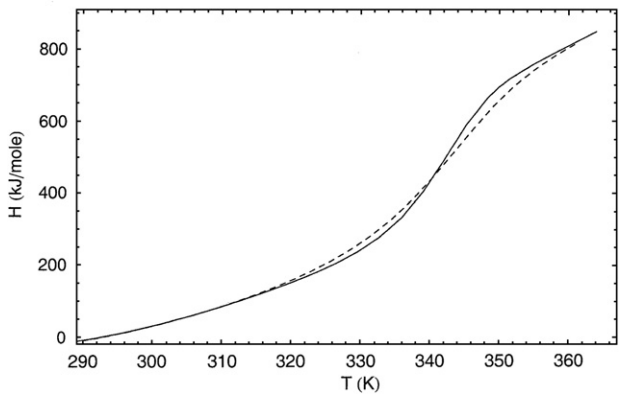

Fig. 11. A comparison of the calculated and empirical enthalpy. The dashed curve gives the calculated enthalpy obtained from Eq. (54) while the solid curve is the empirical enthalpy as given in Fig. 6.

function as given by the following relation

$$H_{\mathrm{m}} = \langle H^{\mathrm{m}} \rangle = \int H^{\mathrm{m}} P(H)\ \mathrm{d}H \tag{56}$$

where $P(H)$ is the enthalpy distribution function giving the probability that the molecule in question has an enthalpy H. We have shown that starting from the heat capacity as a function of temperature

$$C_{\mathrm{p}}(T) = \left(\frac{\partial H}{\partial T}\right)_{\mathrm{p}} \tag{57}$$

we can calculate moments H_{m} defined in Eq. (56) from temperature derivatives of the heat capacity. Specifically, to obtain H_{m} one needs the $(m-2)$th derivative of $C_{\mathrm{p}}(T)$ as is schematically indicated below

$$\frac{\partial^{m-2} C_{\mathrm{p}}}{\partial T^{m-2}} \rightarrow H_{\mathrm{m}} \tag{58}$$

We obtain the temperature derivatives of C_{p} by expanding it as a Taylor series in the temperature as follows

$$C_{\mathrm{p}}(T) = C_{\mathrm{o}} + C_1 \Delta T + C_2 \Delta T^2 + \ldots \tag{59}$$

where

$$\Delta T = T - T_{\mathrm{r}} \tag{60}$$

The quantity T_{r} is the value of the temperature about which the expansion is being made. The coefficients in Eq. (59) are related to the temperature derivatives as follows

$$C_{\mathrm{o}} = C_{\mathrm{p}}(T_{\mathrm{r}}), \quad C_1 = \left(\frac{\partial C_{\mathrm{p}}}{\partial T}\right)_{T_{\mathrm{r}}}, \quad C_2 = \frac{1}{2}\left(\frac{\partial^2 C_{\mathrm{p}}}{\partial T^2}\right) T_{\mathrm{r}} \tag{61}$$

From the relation indicated in Eq. (58), we then have the result that the quadratic expansion of C_{p} given in Eq. (59) contains enough information to calculate the enthalpy moments through H_4.

Given a set of enthalpy moments we then construct the enthalpy distribution function using the maximum-entropy method of Mead and Papanicola [4]. In that approach the distribution function is written in the following form

$$P(H) = \exp[-\phi(H)] \tag{62}$$

where the quantity $\phi(H)$ is a polynomial in H

$$\phi(H) = \sum_{n=0}^{M} \lambda_n H^n \tag{63}$$

and M is the order of the highest moment used. The λ coefficients in Eq. (63) are determined in general by an iteration procedure. The coefficient λ_{o} is determined by the normalization requirement

$$\int P(H)\ \mathrm{d}H = 1 \tag{64}$$

As we have already noted, knowledge of the temperature expansion of $C_{\mathrm{p}}(T)$ through the ΔT^2 term, as indicated in Eq. (58), gives us the first four moments of the enthalpy probability distribution, which, in turn, using the maximum-entropy method, gives the following form for the distribution function

$$P(H) = \exp[-(\lambda_{\mathrm{o}} + \lambda_1 H + \lambda_2 H^2 + \lambda_3 H^3 + \lambda_4 H^4)] \tag{65}$$

The four enthalpy moments determined from the quadratic expansion of the heat capacity allow us to calculate the lambda-parameters given in Eq. (65) from λ_1 to λ_4. Thus basically we are trading knowledge of enthalpy moments obtained from the heat capacity to determine the lambda-parameters in Eq. (65). We note that a $\phi(H)$ function that is quartic in H is capable of exhibiting bimodal behavior which in turn gives a probability distribution with two maxima (representing two significant states).

The key ingredient in constructing enthalpy probability distribution functions is the local expansion of the heat capacity as given by Eq. (59). In Fig. 12 we indicate the values of the temperature (the solid dots) about which we will construct such expansions on the reduced heat capacity curve for the case

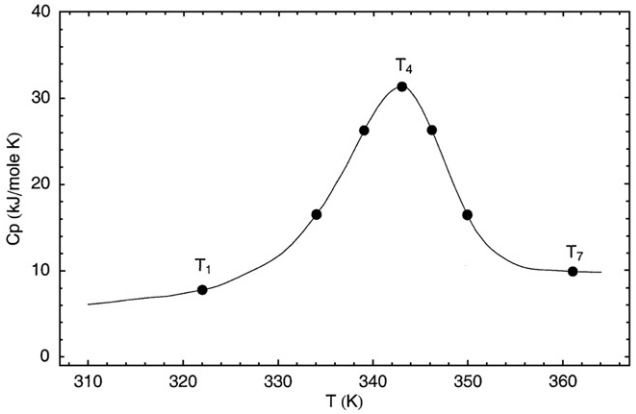

Fig. 12. The reduced heat capacity of the DNA-peptide system given in Fig. 5 showing the set of temperatures (indicated by the solid dots) that will serve as expansion centers for calculating enthalpy probability distributions (from left to right: T=322, 334, 339, 343, 346.15, 349.9, and 361 K).

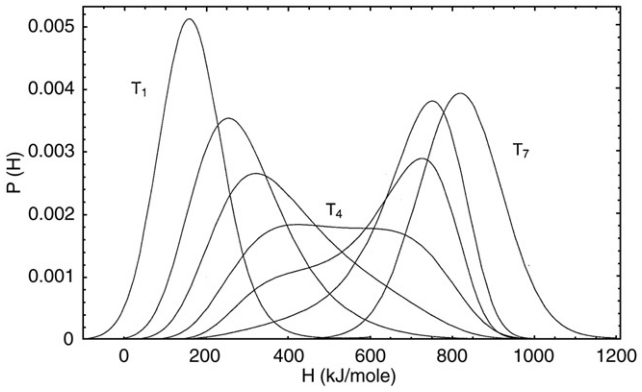

Fig. 13. The enthalpy probability distributions, $P(H)$, given by Eq. (62) for the seven reference temperatures indicated in Fig. 12 and based on the heat capacity curve given in that figure. The temperature values for the curves are, from left to right: T=322, 334, 339, 343, 346.15, 349.9, and 361 K).

r=3.5. (these temperature values are specifically: T=322, 334, 339, 343, 346.15, 349.9, and 361 K). We recall that the reduced heat capacity curve is based on the heat capacity curve labeled complex (r=3.5) in Fig. 1 and was constructed by subtracting out 2.5 times the peptide heat capacity (these represents the background concentration of the peptide). The enthalpy distribution functions for the seven temperature values indicated in Fig. 12 are shown in Fig. 13. One notes that the curves shift to higher values of the enthalpy as the temperature is increased and that the curves go through a maximum in width near the point of the heat capacity maximum.

In Fig. 14 we show the quartic $\phi(H)$ function of Eq. (63) for the three distributions shown in Fig. 13 that correspond to temperatures bracketing the heat capacity maximum (namely, for T=334, 339 and 343 K). The dashed curves represent the net $\phi(H)$ functions obtained from Eq. (62), $\phi(H)=-\ln P(H)$, using the appropriate $P(H)$ curves in Fig. 13. We then decompose the $\phi(H)$ curves into branches that fit the net $\phi(H)$ curves at the low-H and high-H ranges of H, interpolating in the mid-H range. These resulting $\phi(H)$ branches are shown by the solid curves in Fig. 14. In Fig. 15 we convert the $\phi(H)$ functions shown in Fig. 14 into the corresponding $P(H)$ distribution functions using Eq. (62). In Fig. 15 each of the curves has been interpreted in terms of a bimodal distribution (dashed curve) which has been decomposed into two separate distributions (solid curves). The distributions given in Fig. 15 clearly show the transition from mostly complex at 339 K to mostly non-complex at 346 K. Using the relative area under the curves we can calculate the probability of complex and products as a

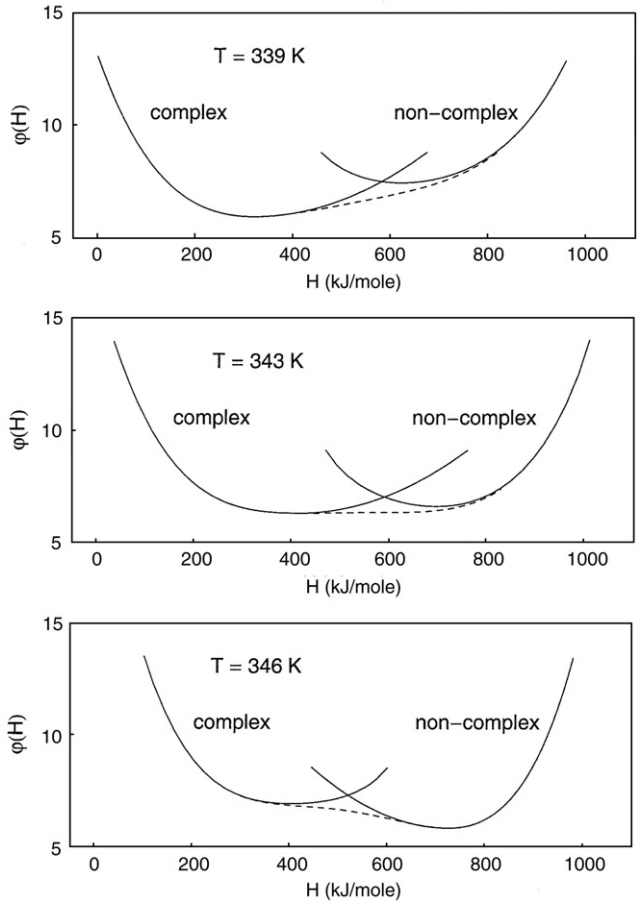

Fig. 14. The decomposition of the net potential, $\phi(H)$, into curves representing the complex and non-complex forms for three values of the temperature.

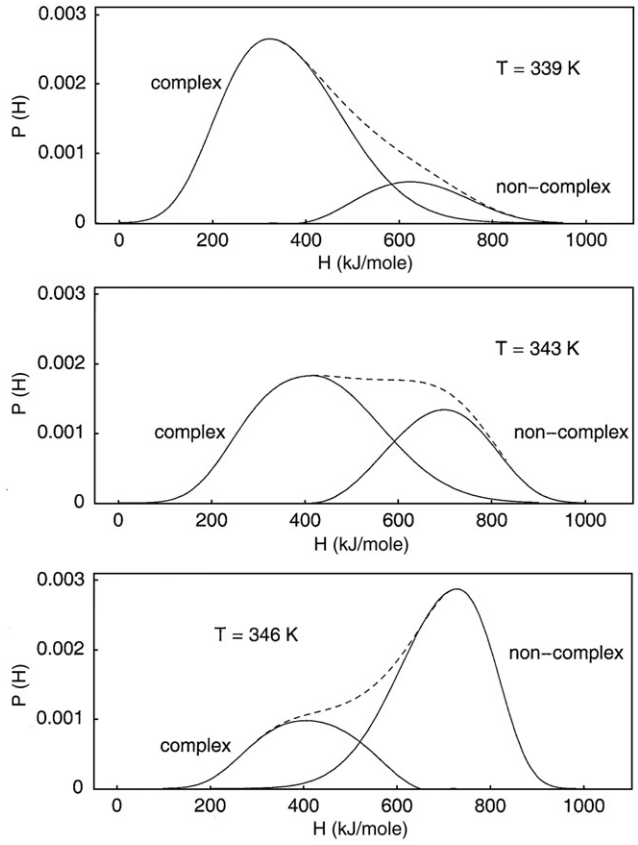

Fig. 15. The enthalpy probability distributions, $P(H)$, from Fig. 13 for the three central values of the temperature. Each of the curves has been interpreted in terms of a bimodal distribution, which is decomposed into two separate distributions. The peaks to the left represent the distribution functions for the intact complex while the peaks to the right give the same quantity for the dissociated state (single-strands of DNA plus peptide).

function of temperature. The results of this calculation are shown in Fig. 16 where the solid dots give the data obtained from the curves in Fig. 15 and two curves not shown. The dashed curve gives the similar quantity calculated using k_3 as given in Eq. (32) and shown (solid curves) in Fig. 10. The two sets of data are clearly describing the same basic process, namely, the conversion of the complex to peptide and single-strands as the temperature is increased.

Finally, we compare the enthalpy differences calculated using the enthalpy-interpolation curves shown in Fig. 6 and the same quantities calculated using the enthalpy probability distributions shown in Fig. 15. Specifically, we will treat the case of $T=346$ K. Given the resolution of the net $P(H)$ curve as shown in Fig. 15, we can then calculate the average enthalpy associated with each curve. The general relation for the average is

$$\langle H_c\rangle = \int H\ P(H)\ \mathrm{d}H \Big/ \int P(H)\ \mathrm{d}H \tag{66}$$

In this manner we obtain the following average enthalpies for the complex and non-complex species:

$$\langle H_c\rangle = 406\ \mathrm{kJ/mol} \quad \text{and} \quad \langle H_{nc}\rangle = 692\ \mathrm{kJ/mol} \tag{67}$$

with the difference in these two quantities given by

$$\langle \Delta H\rangle = \langle H_{nc}\rangle - \langle H_c\rangle = 286\ \mathrm{kJ/mol} \tag{68}$$

These values are indicated by the solid dots in the upper graph in Fig. 17 where the enthalpies are labeled. We note that the average enthalpy associated with the distribution for the complex coincides almost exactly with the position of the maximum in the curve while the same quantity for the distribution for the non-complex does not, reflecting the asymmetry of that curve.

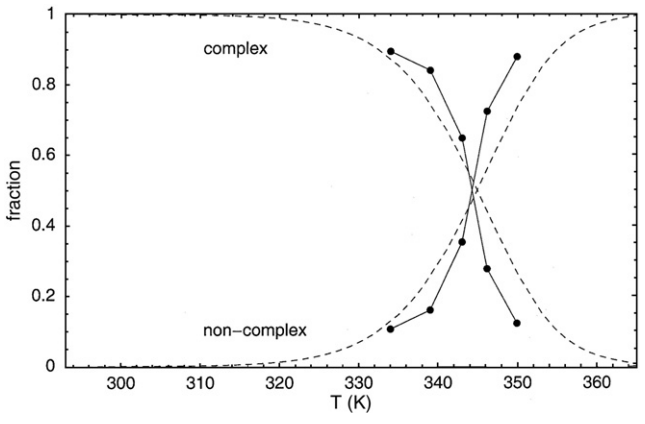

Fig. 16. The extent of reaction-3 as given by Eq. (52) in terms of the variable f (designated here as "fraction"). The solid dots give the results obtained from the $P(H)$ curves shown in Fig. 15 (and from two other cases not shown in Fig. 15) where the probability of intact complex versus dissociated state is determined by the areas under the respective curves in Fig. 15. The dashed curves are reproduced from Fig. 10 and give the same fractions based on k_3 as given by Eq. (32).

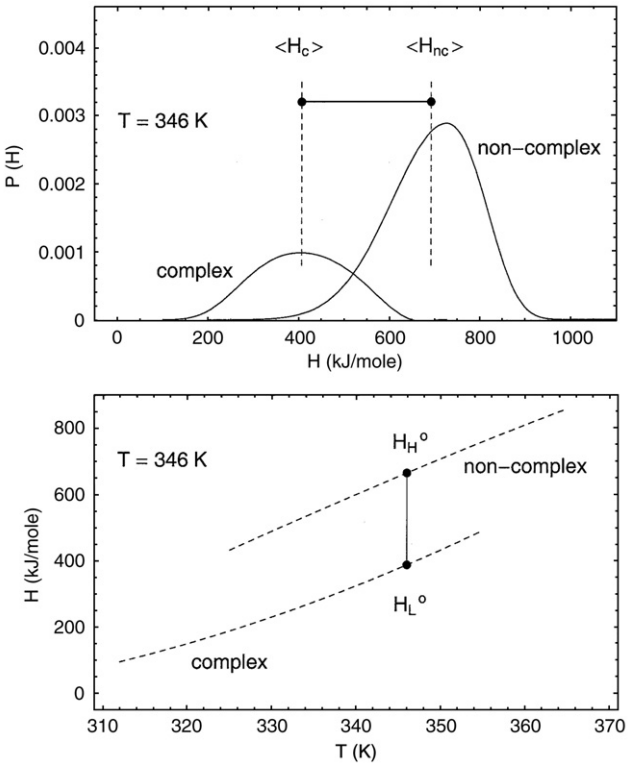

Fig. 17. The upper graph reproduces the $P(H)$ curves given in Fig. 15 for the case of $T=346$ K. The solid dots give the values of the average enthalpy for the complex, $\langle H_c\rangle$, and non-complex, $\langle H_{nc}\rangle$, forms. The lower graph reproduces the two enthalpy branches (dashed curves) shown in Fig. 6, where H_H^o and H_L^o give the enthalpy values at $T=346$ K.

In the lower graph in Fig. 17 we reproduce the two enthalpy branches (dashed curves) shown in Fig. 6. We recall that the lower branch represents the enthalpy as a function of temperature for the complex while the upper branch represents the same quantity for the products (peptide plus single-strands). The solid dots on each branch indicate the appropriate values of H_L^o and H_H^o representative of $T=346$ K. The values of these enthalpies are:

$$H_L^o = 387\ \mathrm{kJ/mol} \quad \text{and} \quad H_H^o = 665\ \mathrm{kJ/mol} \tag{69}$$

with the difference in these two quantities given by

$$\Delta H^o = H_H^o - H_L^o = 277\ \mathrm{kJ/mol} \tag{70}$$

The size of the solid line in each graph gives the value of the enthalpy change for the process complex → products. The values of this quantity obtained from the two approaches illustrated in Fig. 17 are almost identical, that is, to a good approximation:

$$\langle H_c\rangle \approx H_L^o \quad \text{and} \quad \langle H_{nc}\rangle \approx H_H^o \tag{71}$$

6. Conclusion

In this paper we have determined the enthalpy distribution functions for the DNA-peptide system described in Eq. (1) by using the heat capacity data of Dragan et al. [1] as given in Fig. 1.

To understand what species (peptide, DNA single-strands, DNA duplex, DNA-peptide complex) are present as a function of temperature we analyzed the equilibrium properties of this system as outlined in Eqs. (4)–(6). The results of this analysis are given in Figs. 3 and 4, where we find that for the parameter $r=3.5$ (the ratio of initial peptide to DNA concentration) the reaction of Eq. (1) goes cleanly from all-peptide and single-strands to all-complex.

We then used the temperature dependence of the heat capacity to construct the relevant thermodynamic quantities (the enthalpy, entropy and free energy) as given in Figs. 6–9. These quantities are then used to calculate the temperature dependence of the probabilities of complex and non-complex (peptide + single-strands) as shown in Fig. 10. Next we applied the maximum-entropy technique to calculate enthalpy distribution functions for the DNA-peptide system as a function of temperature as shown in Fig. 13. These distribution functions show clear bimodal behavior, indicating two distinct populations – complex and non-complex – as illustrated in Fig. 15.

The two-state nature of the transition from complex to non-complex is illustrated further in Fig. 17 where we show that the ΔH obtained from the bimodal distribution has essentially the same value as the ΔH obtained from the branches of the net enthalpy curves for the complex and non-complex forms.

References

[1] A.I. Dragan, et al., The energetics of specific binding of AT-hooks from HMGA1 to target DNA, J. Mol. Biol. 327 (2003) 393–411.
[2] D. Poland, Enthalpy distribution functions for the unwinding of a short DNA duplex, Biopolymers 81 (2006) 127–135.
[3] D. Poland, Maximum-entropy calculation of energy distributions, J. Chem. Phys. 112 (2000) 6554–6562.
[4] L.R. Mead, et al., Maximum entropy in the problem of moments, J. Math. Phys. 25 (1984) 2404–2417.
[5] D. Poland, Densities of states in gases, liquids and solids, J. Chem. Phys. 113 (2000) 9930–9939.
[6] D. Poland, Ligand-binding distributions in biopolymers, J. Chem. Phys. 113 (2000) 4774–4784.
[7] D. Poland, Ligand binding distributions in nucleic acids, Biopolymers 58 (2001) 477–490.
[8] D. Poland, Enthalpy distributions in proteins, Biopolymers 58 (2001) 89–105.
[9] D. Poland, Distribution functions from moments and the maximum-entropy method, Methods Enzymol. 383 (2004) 427–465.