Forces maintaining the DNA double helix and its complexes with transcription factors

Peter L. Privalov [a], Colyn Crane-Robinson [b, *]

[a] Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA
[b] Biophysics Laboratories, School of Biology, University of Portsmouth, PO1 2DT, UK

ABSTRACT

Precise calorimetric studies of DNA duplexes of various length and composition have revised several long-held beliefs about the forces holding together the double helix and its complexes with the DNA binding domains (DBDs) of transcription factors. Heating DNA results in an initial non-cooperative increase of torsional oscillations in the duplex, leading to cooperative dissociation of its strands accompanied by extensive heat absorption and a significant heat capacity increment. The enthalpy and entropy of duplex dissociation are therefore temperature dependent quantities. When compared at the same temperature the enthalpic and entropic contributions the CG base pair are *less* than that of the AT pair — not more as previously assumed from the extra hydrogen bond. Thus the stabilizing effect of the CG base pair comes from its smaller entropic contribution. The greater enthalpic and entropic contributions of the AT pair result from water fixed by its polar groups in the minor groove of DNA. This water is also responsible for the so-called "nearest-neighbour effects" used to explain the sequence-dependent stabilities of DNA duplexes. Removal of this water by binding DBDs to the minor groove makes this an entropy driven process, in contrast to major groove binding which is enthalpy driven. Analysis of the forces involved in maintaining DNA-DBD complexes shows that specificity of DBD binding is provided by enthalpic interactions, while the electrostatic component that results from counter-ion dispersal is entirely entropic and not sequence-specific. Although the DNA double helix is a rather rigid construction, binding of DBDs to its minor groove often results in considerable DNA bending without the expenditure of significant free energy. This suggests that the rigidity of the DNA duplex comes largely from the water fixed to AT pairs in the minor groove, the loss of which then enables sharp bending.

Contents

* Corresponding author.
 E-mail address: colyn.crane-robinson@port.ac.uk (C. Crane-Robinson).

1. Introduction

Deoxyribonucleic acid (DNA) was discovered in the cell nucleus by Mischer in 1871 and for a long time did not attract much attention as it was assumed that a polymer formed of four simple bases, could play only some supportive, not major, role in contrast to the proteins arranged from 20 different amino acids. Interest in the nucleic acids exploded with the appearance in 1944 of the paper by Avery, McCarty and MacLeod (Avery et al., 1944) showing that the carrier of genetic information is just this DNA. The next essential step was made by Erwin Chargaff who showed that the amount of Adenine in DNA equals that of Thymine, whilst Guanine equals Cytosine (Chargaff, 1950):

[Adenine] = [Thymine] & [Guanine] = [Cytosine]

Based on Chargaff's Rule and the unpublished crystallographic data of Rosalind Franklin showing a twisted helical structure, Watson and Crick proposed an antiparallel double helix structure in which two structurally similar base pairs can form with a purine in one strand paired with its complementary pyrimidine in the opposite strand. Consecutive base pairs stack face-to-face with a spacing of 3.4 Å (Fig. 1).

The double helical structure evoked particular interest because it provided for the coding of genetic information (Gamow and Ycas, 1955). Moreover, the complementary of the strands could explain the mechanism of genetic information replication: the two strands dissociate and new complementary strands are synthesized along each parent template strand.

It was originally supposed that an essential role in maintaining the double helix is played by the hydrogen bonds between the paired bases: two between A and T, three between C and G (Fig. 1). This seemed to be supported by the optical observation that the thermostability of the double helix rises with an increase in the CG content (Marmur and Doty, 1962). This concept was subsequently supported by other studies of DNA stability using various physical methods: their results are summarized in Table 1. According to the three most complete earlier studies the averaged contribution of an AT base pair to the enthalpy of duplex stabilization, is $\Delta H^{AT}=(31 \pm 3)$ kJ/mol-bp and of a CG base pairs $\Delta H^{CG}=(45 \pm 6)$ kJ/mol-bp; correspondingly the averaged entropic contributions amounted to $\Delta S^{AT}=(86 \pm 10)$ J/K·mol-bp and $\Delta S^{CG}=(114 \pm 15)$ J/K·mol-bp. It thus appeared that the extra hydrogen bond between the C and G bases indeed plays a dominant role in supporting the DNA double helix, considerably increasing its enthalpy. It was unclear, however, how to explain its larger entropic contribution, which could not possibly result from its extra hydrogen bond. Understanding the forces involved in double helix maintenance therefore required more direct and detailed investigation of the enthalpy and entropy of forming the DNA double helix.

The next level in understanding genetic information concerns its read-out. This assumes a search for the required sequence along the DNA, followed by a tighter interaction at the target site for further processing of the local information. This is realized by specific proteins, transcription factors, that bind to the DNA using highly specialized DNA binding domains, DBDs. The forces involved in this process cannot be simple: they must be weak enough to permit DBDs to scan easily along the double helix in a search mode and become strong enough for longer-lived binding and, frequently, distortion of the DNA at the target site. Understanding the forces involved in these two phases has required their physical specification in terms of the enthalpic and entropic contributions, determined by direct calorimetric measurements of the heat effects associated with formation of DNA-DBD complexes.

2. Early calorimetry of the DNA double helix

The first calorimetric experiments with highly viscous solutions of polynucleotides used the existing calorimeters for liquids, which were equipped with a mechanical stirrer. The results obtained were not decisive: they just confirmed that with the inclusion of AT base pairs the duplex indeed unfolds at lower temperatures, moreover with a lower heat effect than for CG duplexes (Filimonov, 1980). The insufficiency of this information for understanding the energetic basis of the DNA double helix stimulated the appearance of a more precise calorimetric instrument: the differential scanning microcalorimeter (DSC) for investigating the heat capacity of liquids in a broad temperature range without mechanical stirring (Privalov et al., 1965). The melting profile of phage T_2 DNA obtained with
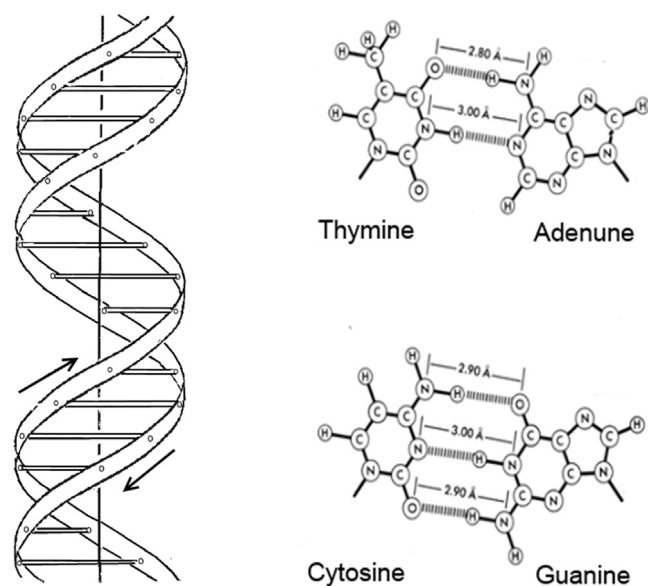

Fig. 1. The DNA double helix and the purine-pyrimidine base pairing proposed by Watson and Crick (1953).

Table 1
Historical thermodynamic data on the contribution of AT and CG base pairs to stabilization of the DNA double helix.

Source	Conditions Temperature	ΔH (kJ/mol-bp)		ΔS (J/K·mol-bp)		ΔC_p (kJ/K·mol-bp)
		AT bp	CG bp	AT bp	CG bp	
Breslauer et al., 1986. Optical, Calorimetry	1 M NaCl 25 °C	33	47	90	113	0
Sugimoto et al., 1996. Optical	1 M NaCl 25 °C	28	46	77	117	0
SantaLucia, 1998. Optical	1 M NaCl 37 °C	33	41	92	113	0
Holbrook et al., 1999. Optical	0.12 M NaCl 25 °C	23		77		0.31
Chalikian et al., 1999. Volumetry, Calorimetry	32 mM NaCl 25 °C	27	29	82	78	0.27

The first three rows give thermodynamic data that have been widely used for the prediction of melting temperatures, T_m. Enthalpy/entropy values were mostly obtained using optical methods and making the assumption of a 2-state (van't Hoff) transition. Breslauer et al. also checked their data calorimetrically by DSC. These authors tabulated their data in terms of the 10 'nearest-neighbour' (NN) model. Three of these ten are all-CG adjacencies and three are all-AT adjacencies: to obtain an approximation of the contribution of an individual CG or AT pair, these two triplet sets have each been separately averaged. The enthalpies and entropies of CG pairs are substantially greater than those of AT pairs in all three data sets. Holbrook et al., 1999 investigated a single 14 bp duplex with no distinction between CG and AT pairs. The data from Chalikian et al. (1999) were obtained using poly[d(AT)]/poly [d(AT)] and poly[d(GC)]/poly[d(GC)] synthetic polynucleotides.

this instrument is presented in Fig. 2. Melting of this DNA proceeded in a rather broad temperature range, so it was impossible to decide if it results in a heat capacity increment and thus to conclude whether the melting enthalpy (ΔH) is temperature dependent (bearing in mind that $\partial \Delta H / \partial T = \Delta Cp$). However, it was found that increase in the salt concentration led to an increase in the DNA melting temperature range and also the heat of melting (Privalov et al. 1969): whilst at 11 mM NaCl the average melting temperature was 64 °C and the enthalpy 36.8 kJ/mol-bp, at 205 mM NaCl the melting temperature rose to 84.8 °C and the enthalpy to 40.3 kJ/mol-bp (see also Krakauer and Sturtevant, 1968). It was unclear, however, if the observed increase in the melting enthalpy resulted from the rise in the melting temperature, i.e. unfolding of the DNA
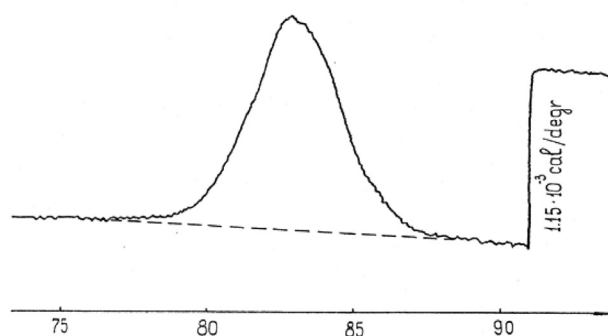

Fig. 2. DSC recording of the heat effect of DNA melting: phage T₂ in 200 mM NaCl, pH 8.5 solution. DNA concentration = 0.75 mg/ml. At the 91 °C point a calibration impulse is shown. (Privalov et al., 1969).

resulted in a heat capacity increment of ~0.17 kJ/K·mol-bp, or it was due to the increasing ionic strength.

The question of the heat capacity increment accompanying DNA unfolding thus remained unsolved: an important issue as this quantity is needed for extrapolating the enthalpies and entropies of unfolding different DNAs from their melting temperatures to the same standard temperature to make them comparable.

The solution required further improvements in the sensitivity, precision and stability of calorimetric instruments, which finally resulted in the appearance of the NanoDSC and also the NanoITC, i.e. an Isothermal Titration Calorimeter. For a review of calorimetric instrumentation see Privalov, 2012.

3. Nano DSC and ITC of the DNA double helix

Fig. 3 panel (**a**) shows Nano-DSC recordings of the heat absorbed upon heating a DNA duplex of 12 CG base pairs arranged in a non-complementary sequence and the heat released on its subsequent cooling. These two heat effects appear as mirror images, showing that the temperature induced unfolding/refolding of this DNA duplex is a highly reversible process.

Correspondingly, the enthalpies of association of the complementary strands of this duplex upon cooling should be of the same magnitude but opposite in sign to the enthalpy of dissociation measured in the heating experiment. According to the usual practice, the heat of DNA melting is determined by extrapolating the initial heat capacity function of the folded duplex to the temperature at which its melting is complete and the disordered complementary strands have separated. It is seen that the linear extrapolation of the initial slope of the heat capacity function (dotted lines) focuses exactly to the heat capacity of the fully unfolded duplex at high temperature. One might conclude, therefore, that duplex unfolding proceeds without any heat capacity increment ($\Delta C_p = 0$), meaning that the enthalpy of the considered reaction does not depend on temperature [$\Delta C_p = (\partial \Delta H / \partial T) = 0$]. Most previous studies led to this conclusion. It appears, therefore, that the excess heat effect of duplex unfolding/dissociation, determined as the area above the extrapolated initial heat capacity line, equals 420 kJ/mol and does not depend on temperature. Thus, the enthalpy of formation of this duplex at room temperature should be of the same magnitude but opposite in sign, −420 kJ/mol.

The right panel (**b**) in Fig. 3 shows an original Nano-ITC recording of the heat effects on titrating one of the strands of the 12 bp CG DNA duplex into its complementary strand at 30 °C. According to this experiment the enthalpy of duplex formation at that temperature is only −160 kJ/mol, a value in sharp contrast to the DSC-measured enthalpy of temperature-induced dissociation, or its association upon subsequent cooling.

There are several possible reasons for the observed discrepancy between the DSC and the ITC measurements of the enthalpy of strand association: (a) the melting enthalpy of the duplex does in fact depend on temperature, i.e. the assumption that DNA melting proceeds without any heat capacity increment is incorrect; (b) the duplex formed at 30 °C is not completely folded; (c) the separated strands are not completely unfolded at 30 °C but have residual structure, so in order to associate they must first unfold this structure and the heat absorbed by this unfolding reduces the observed heat release from duplex formation.

When the melting of this 12 bp duplex is compared with similar 9 bp and 15 bp all-CG duplexes (Fig. 4, upper panel) an increasing molar enthalpy and melting temperature are observed. However, when the observed heats, per base pair rather than per mole, are plotted against the melting temperature (lower panel), a linear temperature dependence of the enthalpies is seen (Inset). This indicates a straightforward additivity of CG base pair contributions to
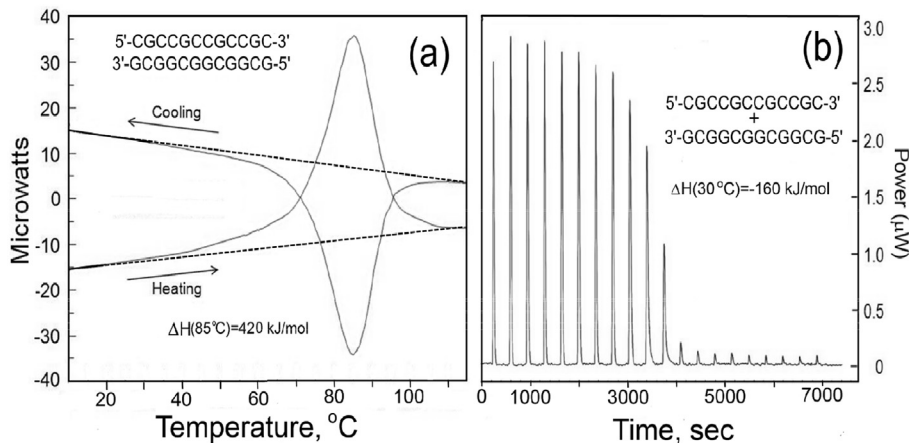
Fig. 3. (a) Original DSC recordings of the heat effect observed upon heating and subsequent cooling of a 12 bp all-CG DNA duplex. (b) The ITC measured heats of forming the 12 bp CG DNA duplex at 30 °C. All measurements in 150 mM NaCl, 5 mM Na-Phosphate, pH7.4 (Reproduced from Vaitiekunas et al., 2015).

the total enthalpy and a heat capacity change of ΔC_p ~0.15 kJ/K.mol-bp. A finite value of ΔC_p means that linear extrapolation of the initial heat capacity function to the transition zone and beyond, as in Figs. 3a and 4, is not justified (Jelesarov et al., 1999; Vaitiecunas et al., 2015). It follows that *the heat capacity function of the duplex at temperatures below the extensive heat absorption peak cannot be simply regarded as the intrinsic heat capacity of fully folded duplex and directly extended to the fully unfolded state at high temperatures*, i.e. the baselines in Figs. 3a and 4 are not justified as they do not take account of the heat capacity increment. It follows that there must be some error in the corresponding enthalpies.

In order to obtain a more accurate estimate of the heat capacity increment, the enthalpy of forming a particular duplex must be made at several different temperatures, not only at the melting temperature, as when using DSC. This can be done using the Nano-ITC, titrating one strand into its complement at several temperatures, as illustrated in Fig. 5, panel (a) for the 12 bp all-CG duplex at 40 °C. The observed association enthalpy at this temperature is only 200 kJ/mol, larger than the 160 kJ/mol measured at 25 °C but much less than the 423 kJ/mol observed by DSC at the T_m of 83.6 °C (Fig. 3).

Moreover, extrapolation of the ITC value at 40 °C to 83.6 °C using a heat capacity increment ΔC_p of 0.15 kJ/K.mol-bp means adding 6.5 kJ/mol-bp, i.e. 78 kJ/mol-duplex, is far from reconciling the two measurements. The large discrepancy results from the presence of residual structure in the separated strands that must first be melted in order to form the complementary duplex in the titration calorimeter (Vesnaver and Breslauer, 1991; Jelesarov et al., 1999; Holbrook et al., 1999; Milev et al., 2003; Vaitiecunas et al., 2015). The heats of melting such residual structures remaining at 40 °C can be determined by heating the individual single strands in the scanning calorimeter: these heats are given by the shaded portions of the DSC scans in panels (c) and (d) of Fig. 5. Despite the complementarity of the two strands they exhibit very different melting profiles: two transitions are seen in the case of the C-rich strand and a single transition with the G-rich strand. It should be noted that residual structures represented by such transitions could be intermolecular as well as intramolecular, a distinction that can be made by monitoring the concentration dependence of the enthalpy and T_m of the transitions. The heats of melting such structures must be added to the ITC—determined magnitude of strand association so as to obtain the heat of forming the duplex from fully disordered single strands. Furthermore, since the duplex is slightly melted at 40 °C (Fig. 5 panel (b)), the small heat released up to this temperature must also be added to the ITC-measured heat so as to obtain the total heat needed to take the duplex from its state at the lowest temperatures up to the fully melted state — the heat which is given by the DSC experiment (for details see Vaitiekunas et al., 2015).

Applying corrections for residual structure to the ITC enthalpies of strand association, the resulting heats can be plotted over a wide temperature range and these are illustrated in Fig. 6 for three all-CG duplexes (Panels (a) and (b)), and for three duplexes of the same length containing centrally located AT pairs (Panels (c) and (d)). From the temperature dependence of these enthalpies it is very evident that *unfolding the DNA proceeds with a heat capacity increment which is identical for duplexes containing only CG pairs and for duplexes of the same length that include AT pairs*, being $\partial \Delta H / \partial T = \Delta C_p = (130 \pm 10)$ J/K·mol-bp, i.e. the heat capacity increment is the same for melting a CG and an AT pair.

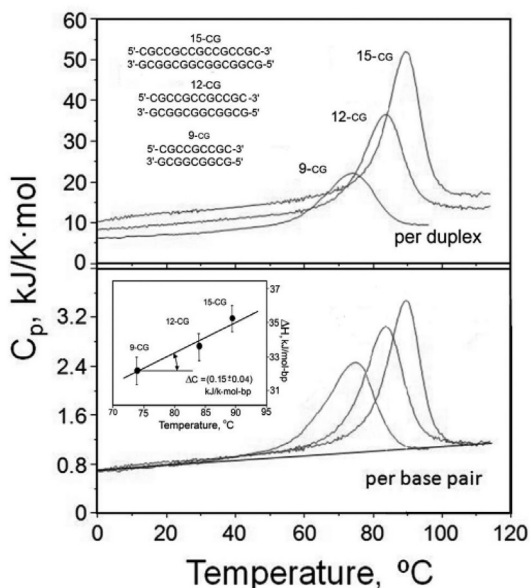**Fig. 4.** The upper panel shows the partial heat capacities of the three CG duplexes in 150 mM NaCl, 5 mM Na-phosphate, pH 7.4, calculated per mole duplex. The lower panel gives the same data calculated per mole base pair. Inset: the dependence of the excess enthalpy, per base pair, on the transition temperature, the slope of which gives the heat capacity increment, (From Vaitiecunas et al., 2015).
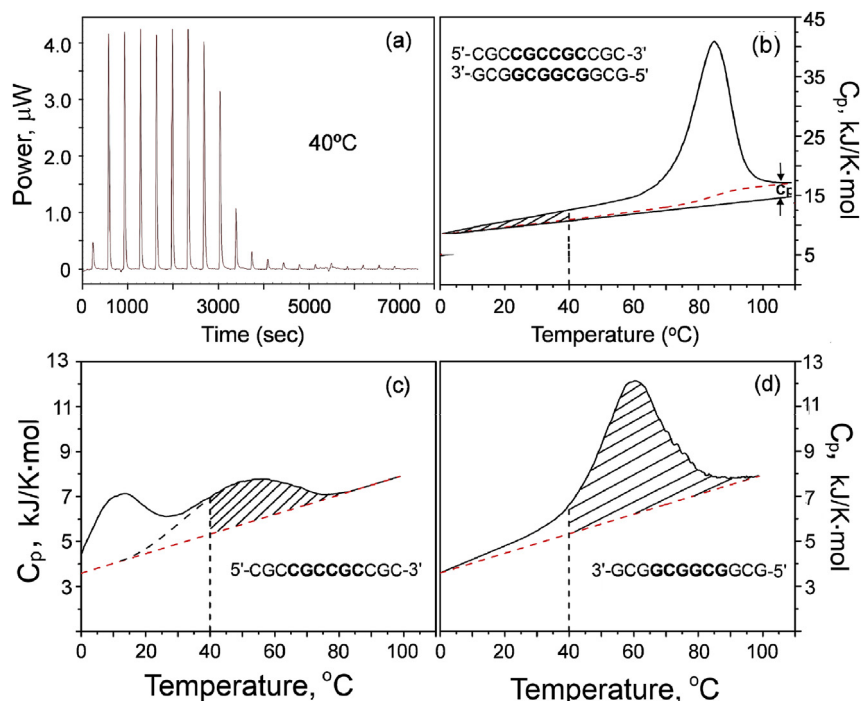
Fig. 5. (a). ITC titration of the 12 bp 5′-CGCCGCCGCCGC-3′ strand with the complementary 3'−GCGGCGGCGGCG−5′ strand at 40 °C. **(b).** The DSC-measured partial molar heat capacity function of the 12-CG duplex; the hatched area represents the enthalpy of the duplex premelting upon heating to 40 °C. **(c)** and **(d)** are the partial heat capacity functions of the two isolated single-strand oligonucleotides; the hatched areas show the enthalpy of the residual structures remaining in the single strands at 40 °C. All in 150 mM NaCl, 5 mM Na-phosphate, pH 7.4. (Reproduced from Vaitiekunas et al., 2015).

This value of ΔC_p is close to that estimated from the DSC enthalpies of Fig. 4 but is more accurate as it is determined from a broader temperature range. It is striking that these ITC enthalpies (coloured dots) − measured between 10 and 45 °C − accurately extrapolate at higher temperatures to the DSC enthalpies measured at their melting temperatures (coloured crosses). This fully reconciles the apparent discrepancies between the ITC and DSC measurements of the melting enthalpies.

Knowledge of the precise value of the heat capacity increment then allows construction of the heat capacity function of a fully folded duplex by subtracting the ΔC_p value from the heat capacity of the fully unfolded duplex at high temperature and connecting this point to the initial heat capacity of the fully folded duplex at low temperatures (Fig. 7). All excess heat effects above this baseline function then represent the total heat effect of duplex unfolding upon heating. Deconvolution analysis of this total excess heat then shows that *the temperature induced change in the DNA duplex is a complex process consisting of two qualitatively different phases: an initial phase* (vertical hatching in Fig. 7) *of gradual energy accumulation − presumably in torsional/stretching vibrations − that intensifies with temperature rise, followed by the phase of cooperative strand dissociation* (horizontal hatching).

As the heat capacity increment takes place only at the phase of cooperative dissociation of strands, determination of the total enthalpy of duplex dissociation at some standard temperature (a necessity for comparing various duplexes) requires extrapolating the enthalpy of the *cooperative phase* to this temperature (using the observed heat capacity increment) and summing this with the heat of the *gradual phase* − which does not depend on temperature as the heat capacity increment of this phase is zero. The total enthalpy thereby determined from the DSC experiment corresponds in magnitude to the ITC measured enthalpy of duplex association when the latter is corrected for the contribution of residual structures in the separated strands as shown in Fig. 5 (Vaitiecunas et al., 2015). We therefore concentrate on the cooperative phase of DNA strand separation because *the gradual phase, that proceeds without a heat capacity increment, does not contribute to the Gibbs energy of double helix stabilization.*
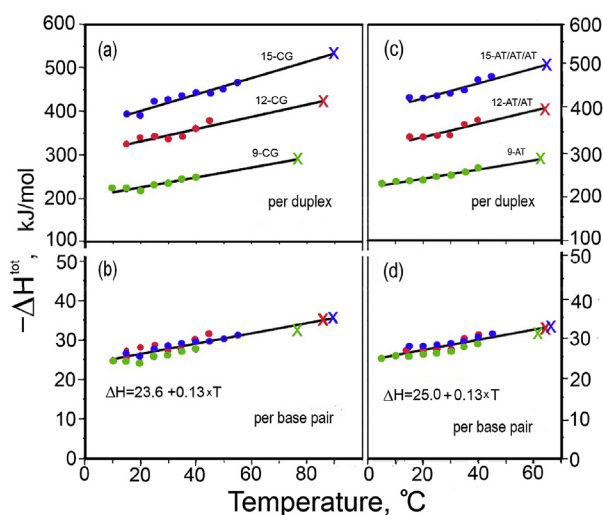
Fig. 6. Panels (a) and (c) are the ITC-measured molar enthalpies of formation of the 9-CG, 12-CG and 15-CG duplexes (of Fig. 4) and 9-AT (CGCAAACGC); 12-AT/TA (CGCAAATTTCGC) and 15-AT/TA/AT (CGCAAATTTAAACGC) duplexes corrected for residual structures in the separated strands and for premelting. (b) and (d) are the specific molar (i.e. per base pair) enthalpies of these duplexes. Crosses indicate the total enthalpy values of formation of the considered duplexes obtained from the DSC-measured excess heat of duplex melting and attributed to the transition temperatures, T_t. All in 150 mM NaCl, 5 mM Na-phosphate, pH 7.4 solutions (reproduced from Vaitiecunas et al., 2015).

Fig. 7. The observed heat capacity profile of the 12-CG duplex. The expected heat capacity of the fully folded DNA duplex is obtained by subtracting the heat capacity increment, ΔC_p, from the heat capacity of DNA at 100 °C and linearly extrapolating back to the start of melting at 0 °C. The experimental excess heat effect is then deconvoluted into the non-cooperative (gradual, vertical hatching) and cooperative (horizontal hatching) phases.

4. Enthalpic contribution of the base pairs to DNA duplex maintenance

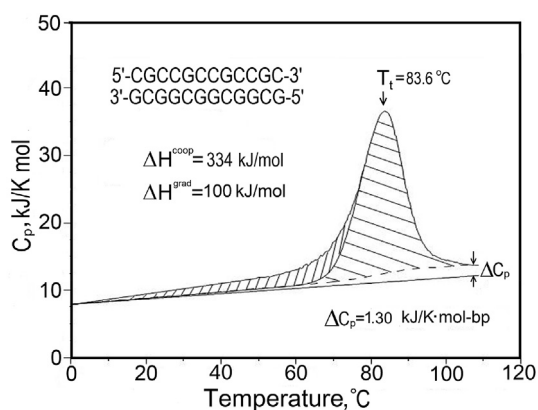Although the substitution of CG base pairs by AT pairs results in a *decrease* of the thermal stability (melting temperature) of the duplex, absolutely unexpected was observing an *increase* in the enthalpy of duplex unfolding. The ITC data in Fig. 6 show that when corrected to the same temperature the total enthalpy of forming a duplex containing AT base pairs is somewhat larger than those of a duplex of the same length consisting only of CG base pairs. This becomes strikingly obvious on comparing the DSC measured heats of melting the three all-CG duplexes with three of the same length containing increasing numbers of AT pairs (Fig. 8).

It can be seen that with increasing numbers of AT pairs the lower melting peak becomes greater in area than the all-CG peak, i.e. the enthalpy of melting an AT pair is greater than for a CG pair. Moreover, if the enthalpies are corrected to identical temperatures using the known value of ΔC_p, the enthalpy difference between the duplexes of identical length increases. Overall, it is clear that *the enthalpic contribution of the AT base pair definitely exceeds that of the CG base pair.* If the enthalpy of AT dissociation is significantly larger than that of CG dissociation, it immediately follows that the *entropy of AT melting must be very significantly greater than that of CG pairs, as duplexes containing AT pairs melt at lower temperatures.*

The main conclusion of the very unexpected enthalpic and entropic contributions of the base pairs is that *the CG-rich DNA duplex is more stable than the AT-rich duplex not because the enthalpy of CG base pair dissociation is larger than that of AT but because the entropy of its dissociation is lower.* Alternatively this could be stated as: *the AT-rich duplex is less stable than the CG-rich duplex because the entropy of AT dissociation is larger than the entropy of CG dissociation.*

What then is the precise enthalpic and entropic contributions of the AT base pairs in duplexes containing AT runs? They can be obtained by first extrapolating the measured enthalpies of the AT-containing duplexes to the standard temperature of 25 °C using $\Delta C_p = (130 \pm 10)$ J/K·mol-bp. The enthalpy of the smallest, 9 bp AT duplex, consisting of 6CG and 3AT base pairs, was then subtracted from the enthalpy of each of the longer AT-containing duplexes and the result divided by the difference in the number of AT pairs, giving the enthalpic contribution of a single AT pair at 25 °C, i.e. an averaged value without regard for its neighbours. This showed that
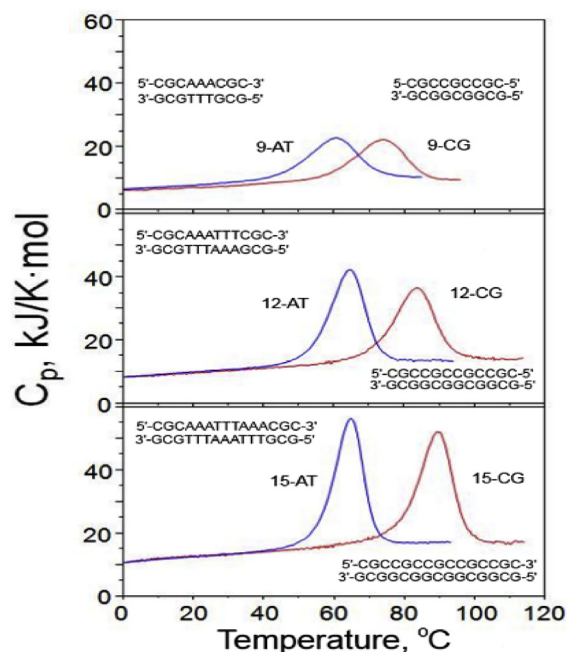

Fig. 8. Comparison of the partial molar heat capacities of the 9, 12 and 15 base pair all-CG duplexes (in red) and the same length duplexes having AT pairs in the central region (in blue). These are the same 6 duplexes shown in Fig. 6. All measurements at the identical concentration of 283 μM in 150 mM NaCl, 5 mM Na-Phosphate, pH 7.4. (from Vaitiekunas et al., 2015).

while the enthalpy of dissociating a CG pair is (18.8 ± 0.3) kJ/mol-bp, for an AT pair the enthalpy is (25.0 ± 3.0) kJ/mol-bp (see Table 2). Thus, in contrast to all previous publications *the enthalpic contribution of the AT base pair is ~30% larger than that of a CG pair at the standard temperature of 25⁰C.*

5. Entropic contribution of the base pairs to DNA duplex stabilization

The situation with the entropic contribution of the AT and CG base pairs is significantly more complicated than their enthalpic contributions. As shown in Fig. 4 the enthalpy of the DNA duplex increases in proportion to its length and as the stability of the duplexes rises with increase in the number of base pairs it follows that the entropy of unfolding is not proportional to the length. This is because the total duplex unfolding entropy — that increases with the number of bases and depends on temperature — also includes an entropy term that results from the appearance of a new kinetic unit on dissociation of the duplex, which does not depend on the number of bases or the salt conditions or the temperature. This entropy is usually regarded as a *translational entropy*.

According to the view originally proposed by Gurney (1953), the translational entropy is expressed by the cratic term, δS^{cratic}, which is just the entropy of mixing the additional kinetic unit into the solvent, following dissociation of a complex and is assumed to be independent of the solution composition. For 1 M standard aqueous solution (containing 55 mol of water) $\delta S^{cratic} = R\ln(1/55) = 8.03$ cal/K·mol $= 34.5$ J/K·mol for dissociation of a dimer, and is supposed to be independent of the molecular weight of the solute. However, the cratic entropy became a target of severe criticism from theoreticians as physically ungrounded: based on the statistical mechanics of gases, it was suggested that values of the translation-rotational entropy should be of the order of 400 J/K·mol (Finkelstein and Janin

Table 2
Optimized contributions of the CG and AT base pairs to the enthalpy, entropy and heat capacity increment of double helix dissociation at 25 °C. Also given are the magnitudes of the translational entropy increase on helix dissociation, which equates to the stoichiometric correction coefficient. Data from the present analysis and from Privalov and Crane-Robinson (2018), all obtained at a duplex concentration of 283 μM in 150 mM NaCl, 5 mM Na Phosphate pH 7.4 .

Base pair	ΔCp kJ/K·mol-bp	ΔH^{coop} kJ/mol-bp	ΔS^{conf} J/K·mol-bp	$\Delta S^{trans} = R \ln\left(\frac{2}{[No]}\right)$	
				$[N_o]$ μM	ΔS^{trans} J/Kmol
CG		18.8 ± 0.3	44.7 ± 0.5	403	70.7
				337	72.2
	0.13 ± 0.01			283	73.6
				214	76.0
AT		25.0 ± 2.0	72 ± 5.0	107	81.7
				40	90.0

(1989), Janin, (1995). Very similar values for the entropy effects of dimer dissociation were obtained by Tidor and Karplus (1994) using the statistical-thermodynamic approach of Chandler and Pratt (1976). According to these authors, dimerization of insulin should result in a decrease of the translational entropy by 180 J/K·mol and a decrease of the rotational entropy by 200 J/K·mol, but this should be accompanied by an increase of the vibrational entropy by 110 J/K·mol; thus the overall change in the external entropy (i.e. the entropy not associated with changes in conformation or hydration) upon dimerization of insulin should amount to $\Delta S^{trans} = -270$ J/K·mol. Translational entropy values in the range from 300 to 400 J/K·mol have been widely used by many authors in the thermodynamic analysis of the formation of protein/protein and protein/DNA complexes (see e.g. Janin & Chothia (1990); Janin (1995); Spolar & Record (1994); Searle et al. (1992). However, early calorimetric studies of unfolding an S-S crosslinked and non-crosslinked dimeric globular protein and also an α−helical coiled-coil in aqueous solution showed that the translational entropy increase is much lower than suggested by the statistical mechanics of gases (Tamura and Privalov, 1997; Privalov and Tamura, 1999; Yu et al., 1999).

As the enthalpies and therefore the conformational entropies of the all-CG duplexes are strictly linear with length, they are suitable objects for extracting the translational entropy from the total entropies. From calorimetric data, at 283 μM concentration the 15 and 9 base-pair CG duplexes unfold cooperatively at temperatures 362.7 K and 347.2 K, respectively, with enthalpies of 408 kJ/mol and 223 kJ/mol (for details see Vaitiecunas et al., 2015).

The total entropies of unfolding at their transition temperatures are then:

$$\Delta S_t^{tot}(15CG) = \frac{408kJ/mol}{362.7K} = 1125 J/K \cdot mol \tag{1}$$

and

$$\Delta S_t^{tot}(9CG) = \frac{223kJ/mol}{347.2K} = 642 J/K \cdot mol \tag{2}$$

Extrapolating these entropies to the standard temperature of 25 °C (using $\Delta Cp = 0.13$ kJ/K·mol-bp) and expressing the total entropy as the sum of the conformational and translational components, we have:

$$743 \text{ J/K·mol} = \Delta S_{25}^{15CG(conf)} + \Delta S^{trans} \quad \text{and}$$
$$464 \text{ J/K·mol} = \Delta S_{25}^{9CG(conf)} + \Delta S^{trans} \tag{3}$$

As both experiments were carried at the same duplex concentration, i.e. the translational entropies are the same in each case and the conformational entropies are additive, subtracting one from the other and dividing by the difference in the number of base pairs, at the duplex concentration of 283 μM one gets:

$$\Delta_9^{15} S_{25}^{conf}/6 = (46.5 \pm 3.0) \text{J/K·mol-bp}. \tag{4}$$

Using this value of the conformational entropy of a CG pair, the translational entropy is best evaluated by analyzing the dependence of duplex thermostability (i.e. the melting temperature, T_t) on the number of base pairs. Bearing in mind that the heat capacity increment upon duplex dissociation, ΔC_p, does not depend on temperature, the duplex transition temperature can be expressed by the straightforward equation:

$$T_t = \frac{\Delta H_t^{CG}}{\Delta S_t^{CG}} = \frac{\left[\Delta H_{25}^{CG} + \Delta C_p\left(T_t^1 - 298\right)\right] \times n^{CG}}{\left[\Delta S_{25}^{CG} + \Delta C_p \ln\left(T_t/298\right)\right] \times n^{CG} + \Delta S^{trans}} \tag{5}$$

From this equation, we have for the translation entropy, ΔS^{trans}:

$$\Delta S^{trans} = \left[\Delta H_{25}^{CG} + \Delta C_p \times (T_t - 298) \times n^{CG}\right]/T_t - \left[\Delta S_{25}^{CG} + \Delta C_p \times \ln\left(\frac{T_t}{298}\right)\right] \times n^{CG} \tag{6}$$

The derived values of ΔS^{trans} are very sensitive to the magnitude of the conformational entropy, ΔS_{25}^{CG}, which carries some significant experimental error − but the translational entropy should *not* depend on the number of base pairs in the duplexes, nor on the conformational entropy of the bases. This requirement is realized at a conformational entropy value of 44.6 J/K·mol-bp, i.e. within the above experimental error, and for which the translational entropy is calculated to be $\Delta S^{trans} = (73.2 \pm 0.5)$ J/K·mol at 283 μM duplex concentration. This analysis therefore optimizes both the conformational entropy of a CG pair and also the translational entropy of the duplex (see Privalov and Crane-Robinson 2018 for more details).

It is notable that the translational entropy thus obtained for separation of the DNA strands is at least five times smaller than that derived by statistical mechanics for the dissociation of dimeric macromolecules in the gas phase (Finkelstein and Janin (1989); Janin, 1995; Tidor and Karplus (1994); Chandler and Pratt (1976); Janin & Chothia (1990)) and also differs from the cratic entropy value proposed by Gurney (1953). However, it is essentially identical to the stoichiometric correction term for a multimeric reaction, such as:

$$N \leftrightarrow m_1 D_1 + m_2 D_2 + m_3 D_3 + \dots + m_k D_k, \tag{7}$$
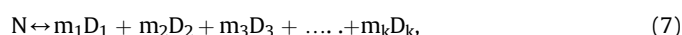
The equilibrium constant for this reaction at the transition midpoint (F = 1/2) is expressed as (Privalov, 2012):

$$K = ([N]_o/2)^{n-1} \prod_{i=1}^{k} [m_i]^{n_i} \tag{8}$$

where $n = \Sigma m_i$ is the order of reaction. Correspondingly,

$$\Delta G_t = -RT_t(n-1)\ln([N]_o/2) \prod_{i=1}^{k} [m_i]^{m_i} \tag{9}$$

For the case of a heterodimer, e.g. $N \Leftrightarrow D_1 + D_2$, when $m_1 = 1$, $m_2 = 1$, $n = 2$, $K = [N_o]/2$:

$$\Delta G_t = -RT_t \ln([N_o]/2) = \Delta H_t - T\Delta S_t \qquad (10)$$

so:

$$\Delta S_t = \frac{\Delta H_t - \Delta G_t}{T_t} = \frac{\Delta H_t}{T_t} + R\ln([N_o]/2) = \frac{\Delta H_t}{T_t} - R\ln(2/[N_O]) \qquad (11)$$

where $[N_o] = N_o/N_{st}$ is the dimensionless initial concentration of the complex and $R = 8.31$ J/K·mol is the universal gas constant. Here $\Delta H_t/T_t$ is the whole entropy of duplex dissociation at T_t, which comprises two components: the conformational, ΔS_t and the correction term $R\ln([N_o]/2)$. At the DNA duplex concentration of 283 μM used in our experiments, the term $R\ln([N_o]/2)$ amounts to 73.6 J/K·mol. It is clear therefore that *the translation entropy is fully expressed by the stoichiometric correction term.*

Although the translational contribution to the total entropy at 25 °C appears to be small, it is *totally* responsible for the difference in the melting temperatures of, for example, the 9 and 15 bp all-CG duplexes. The constant contribution of the translational entropy is relatively large for a duplex as short as 9 base pairs but as it grows in length ΔS^{trans} becomes a decreasing proportion of the total entropy. Data for the translational entropy are included in Table 2 together with the conformational entropies and enthalpies. Most notable are the large differences between the enthalpic and particularly the entropic contributions of the CG and AT base pairs: the difference between the entropic contributions of the two base pairs amounts to 27 J/K·mol-bp, i.e. at 25 °C the difference in the entropy factors, TΔS, is 8 kJ/mol-bp.

6. Prediction of DNA stability from the thermodynamic data

The single-valued enthalpies and entropies of CG and AT pairs given in Table 2, together with the translational entropy, can be used for predicting the stability of DNA duplexes of various composition, length and concentration. The method of doing this differs from current protocols that assume the enthalpy and entropy are independent of temperature (Breslauer et al., 1986; Sugimoto et al., 1996; SantaLucia, 1998). The present data, at 25 °C, must first be corrected to some 'expected' T_m, say 348 K, using the known magnitude of ΔCp, and the enthalpy divided by the entropy to obtain a predicted T_m — after including the translational component in the total entropy — using an equation analogous to (5) but also including AT pairs:

$$T_m = \frac{\Delta H}{\Delta S}$$
$$= \frac{[\Delta H^{AT} + \Delta C_p \times (348-298)]n^{AT} + [\Delta H^{CG} + \Delta C_p \times (348-298)]n^{CG}}{[\Delta S^{AT} + \Delta C_p \ln(\frac{348}{298})]n^{AT} + [\Delta S^{CG} + \Delta C_p \ln(\frac{348}{298})]n^{CG} + \Delta S^{trans}} \qquad (12)$$

This predicted T_m is then used in further iterations. The full procedure is given in detail in Privalov and Crane-Robinson (2018) and the results are presented in Table 3. For the all-CG duplexes the final predicted T_m values correspond to the observed within ±0.3 K, which is only ±0.1% on the absolute temperature scale. For example, T_m for the 15 bp all-CG duplex is predicted to be 89.5 °C, while the experimental value is 89.3 °C. For duplexes containing AT base pairs the correspondence is an order of magnitude less accurate (±1.5 K, i.e. ±0.5%) and one can see that the values of T_m depend not only on the proportion of AT pairs but also their arrangement along the duplex. This effect of base pair sequence on DNA duplex stability was first noticed by Tinoco and coworkers (Borer et al.,

Table 3
The melting temperatures of various DNA duplexes calculated using the data given in Table 2 for CG and AT pairs and the iterative use of equation (12) (see also Privalov and Crane-Robinson, 2017).

DNA duplexes	Composition	Conc. μM	T_t °C Exp.	T_t °C Calc.
5'–CGCCGCCGC–3' 3'–GCGGCGGCG–5'	9-CG	283	74.0	73.5
5'–CGCCGCCGCCGC–3' 3'–GCGGCGGCGGCG–5'	12-CG	283	83.6	83.1
5'–CGCCGCCGCCGCCGC–3' 3'–GCGGCGGCGGCGGCG–5'	15-CG	283	89.5	89.3
5'–CGCCGCCGCCGCCGCCGCCG–3' 3'–GCGGCGGCGGCGGCGGCGGC–5'	20-CG	283	N/A	94.2
5'–CGCCGCCGCCGCCGCCGCCGCGCGC–3' 3'–GCGGCGGCGGCGGCGGCGGCGCGCG–5'	25-CG	283	N/A	98.7
5'–CGC**AAA**CGC–3' 3'–GCG**TTT**GCG–5'	6CG, 3AT	283	60.4	61.1
5'–CGC**AAATTT**CGC–3' 3'–GCG**TTTAAA**GCG–5'	6CG, 6AT	283	64.5	62.3
5'–CGC**AAAAAA**CGC–3' 3'–GCG**TTTTTT**GCG–5'	6CG, 6AT	283	63.0	62.1
5'–CGC**TTTAAA**CGC–3' 3'–GCG**AAATTT**GCG–5'	6CG, 6AT	283	60.8	62.1
5'–CGC**ATATAT**CGC–3' 3'–GCG**TATATA**GCG–5'	6CG, 6AT	283	60.3	62.1
5'–CGC**AAATTTAAA**CGC–3' 3'–GCG**TTTAAATTT**GCG–5'	6CG, 9AT	283	64.8	63.3
5'–CGC**AAAAAAAAA**CGC–3' 3'–GCG**TTTTTTTTT**GCG–5'	6CG, 9AT	283	65.1	63.3
5'–CGC**AGAGAGAGAG**CGC–3' 3'–GCG**TCTCTCTCTC**GCG–5'	10CG, 5AT	283	72.2	73.6
5'–CGC**ACACACACAC**GCG–3' 3'–GCG**TGTGTGTGTG**CGC–5'	10CG, 5AT	283	75.9	73.6
5'–GCG**AACAAT**CGG–3' 3'–CGC**TTGTTA**GCC–5'	7CG, 5AT	403	64.8	67.3
5'–GCG**AACAAT**CGG–3' 3'–CGC**TTGTTA**GCC–5'	7CG, 5AT	87	60.6	61.5

1974) and is described in the literature as a consequence of Nearest-Neighbour (NN) Interactions. *It is now clear that this effect depends on the presence of the AT base pair.* The nature of these interactions, however, has so far been absolutely obscure.

There have been many attempts to take NN Interactions into account by monitoring how the enthalpic and entropic contributions of various combinations of base pairs depend on their mutual disposition (Breslauer et al., 1986; SantaLucia, 1998; Allawi and SantaLucia, 1997; Sugimoto et al., 1996). For example, according to Breslauer et al. (1986), while the enthalpic contribution of the AT/TA adjacency is 8.6 kcal/mol, for TA/AT it is 6.0 kcal/mol; correspondingly their entropic contributions were 23.9 and 16.9 cal/K·mol; for the CG/GC and GC/CG adjacencies the enthalpic contributions were 11.9 and 11.1 kcal/mol and their entropies 27.8 and 26.7 cal/K,mol, correspondingly. A similar approach used by other authors led to good correspondence between the predicted and measured thermostabilities of the DNA duplexes. It is striking, however, that the enthalpic and entropic contributions of the CG base pairs in all these models exceeded those of the AT, in accord with the widely held beliefs of those times. However, our recent studies with the most precise calorimetric instrumentation showed that this is not the case (Table 2): both the enthalpic and entropic contributions of the AT base pair significantly exceed those of CG. Thus, the DNA stabilizing effect of the CG base pair results not from its larger enthalpic contribution but from its smaller entropic contribution than that of the AT pair. The question is then: *why is the entropic contribution of the AT base pair is so large?*

7. The role of water in stabilization of the DNA duplex

The only possible explanation for the observed large excess of the enthalpy and particularly the entropy of AT base pairing over

that of CG *is water fixed by the AT pair in the minor groove of DNA* (Fig. 9A). This was first noticed crystallographically (Drew and Dickerson, 1981) showing that water is fixed by the N3 of A and O2 of T groups of the AT pair (Kopka et al., 1983; Prive et al. 1987). NMR studies (Liepinsh et al., 1992; Johannesson and Halle, 1998) also revealed this immobilized water.

In addition to this primary layer of fixed water molecules (blue in 9A) runs a secondary layer of waters (yellow in 9A) donating H-bonds to the primary shell of oxygen atoms that thereby assume the tetrahedral coordination specific for ice (Shui et al., 1998; Arai et al., 2005). Further evidence indicates that an 'outer spine' of third and fourth-shell water molecules forms a pattern of fused hexagons (Egli et al., 1998; Tereshko et al., 1999). High resolution crystallography has also shown (Prive et al., 1987) the presence of two arrays of water molecules that bridge between purine N3 and pyrimidine O2 atoms to the O4's of adjacent sugar rings, providing a regular lining to both walls of the minor groove (Fig. 9B). Release of tightly bound minor groove water into the bulk solution will result in positive contributions to both the enthalpy and entropy of melting and since the 'ice-like' water (Fig. 9A) is bound to AT pairs it will augment their relative contributions − as seen in Table 2.

It is striking that the excess entropy contribution of an AT pair, relative to CG, is 27 J/K·mol (Table 2), which is higher than the entropy of melting ice, 22 J/K·mol. This indicates that a water molecule fixed by an AT pair also affects the state of neighbouring water molecules, i.e. *the AT pair clusters neighbouring water molecules in the minor groove of DNA*. The water molecules fixed by AT pairs in the minor groove might be the cause of the Nearest-Neighbour effect. Investigation of possible interactions between fixed water clusters in the minor groove of DNA by computer modeling is of significant interest for understanding the energetic basis of the DNA double helix.

The other key feature of the DNA duplex is the significant heat capacity increment that appears on its dissociation. This certainly could not result from the increase of conformational freedom of the DNA strands upon their dissociation but it might be caused by the water bound/released in this process. If so, then why is this heat capacity increment essentially identical for the AT and CG base pairs, despite the drastic difference in their influence on the state of water in the minor groove? Notably, the same heat capacity increment of 130 J/K·mol-bp determined for AT and CG deoxy-nucleotide pairs was also measured for the poly(A)poly(U) ribo-nucleotide double helix (Filimonov and Privalov, 1978). It would appear that there is some other source of the heat capacity increment.

The increase in the heat capacity on DNA duplex dissociation is not in fact surprising: DNA strand separation results in exposure of the very similar apolar surfaces of the AT and CG bases pairs to water and it is well known that the hydration of apolar groups results in an increase of their partial heat capacities (for reviews see: Kauzmann, 1959; Privalov and Gill 1988; Makhatadze and Privalov, 1995). It remains surprising, however, that despite the known heat capacity effect of apolar group hydration and the understanding that dissociation of a duplex results in exposure of the nonpolar surfaces of the bases to water, the enthalpy of DNA unfolding was overwhelmingly believed to be temperature independent.

8. The interactions of DNA with transcription factors

Transcription factors recognize certain sequences in DNA by means of specialized DNA-binding domains, DBDs, that bind to target sites, sometimes deforming the double helix for further processing of the encoded genetic information. The various DBDs differ considerably in their structure and in their manner of interaction with the target DNA. Fig. 10 illustrates the DNA binding of four very different types of sequence-specific DBD.

The first question that arises in considering these very different DBDs is whether they have some common physical feature characteristic of the family? One such feature is their specific heat capacity function that increases steeply from the very beginning of heating − in contrast to proteins defined as globular, i.e. that have a
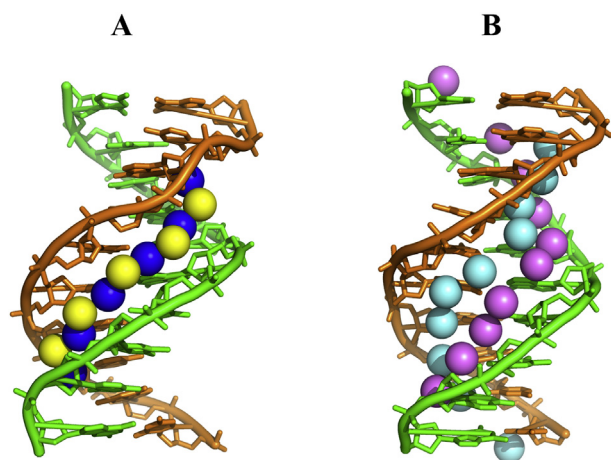
Fig. 9. (A) Display of primary (*blue*) and secondary (*yellow*) layers of the spine of water in the minor groove of the crosslinked dodecamer CGCGAATTCGCG, generated from the coordinates of NDB accession number BD0008 [reproduced from Privalov et al. (2007)]. (B) Crystal B-form structure of the mixed sequence decamer CCAAGATTGG showing strings of water molecules observed along the walls of the minor groove (cyan and magenta spheres). Water molecules bridge from purine N3 or pyrimidine O2 atoms to the O4′ atoms of adjacent deoxyribose rings. [Prive et al., 1987; 3DNB].
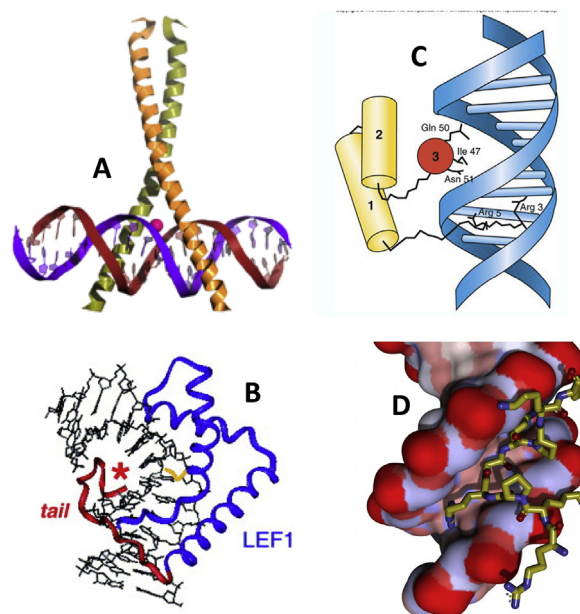
Fig. 10. Examples of DNA complexes with DBDs from various transcription factors: (A) GCN4-**b**ZIP. A homodimer of two continuous α−helices, each comprising a **b**asic region contacting the major groove, plus the leucine **zip**per dimerization region. (B) HMG Box from mLEF1. This includes the minimal Box (blue) binding to the minor groove of the bent DNA, plus the basic C-tail (red) that crosses the major groove on the inside of the bend∗. A DNA bend of >90° is generated. (C) Drosophila Antennapedia HTH homeo-domain: helix 3 (red) is the recognition element bound in the major groove while the N-terminal tail extends into the minor groove. (D) The second AT-hook (PRGRP) from HMGA1 lying along the base of the minor groove in AT-rich DNA.

fully compact and rigid structure. This is illustrated in Fig. 11 for the HMG box from Sox5 and the GCN4-bZIP homo-dimer.

Among globular proteins, the most compact and stable is bovine pancreatic trypsin inhibitor (BPTI) that is heavily S-S crosslinked: upon heating in aqueous solution it unfolds above 100 °C (Makhatadze et al., 1993). Fig. 12 compares the low temperature heat capacity functions of a range of proteins, both globular and DBDs. The heat capacity of BPTI (N2 in Fig. 12) only slightly exceeds that of dry protein (N1), which never unfolds upon heating. In contrast, the heat capacities of all DBDs (N11 to N16) increase steeply with temperature, showing that they intensively fluctuate at low temperatures.

Compact globular proteins consisting of a single domain, e.g. lysozyme and myoglobin, have a single hydrophobic core and melt as a unitary cooperative unit, i.e. show single peak heat capacity functions (for reviews see Makhatadze and Privalov, 1995; Privalov, 2007). In contrast, deconvolution analysis of the heat capacity functions of several HMG boxes (Fig. 13, Lef and Sry) shows that they do not represent single transitions but comprise two or three distinct stages upon heating (Dragan et al., 2004b; Privalov and Dragan, 2007). This shows that at lower temperatures, below physiological, they are already partially denatured and therefore exhibit increased heat capacities.

What then happens if the rather loose structures formed by the free HMG boxes bind to their target DNA? Fig. 14(a) shows that
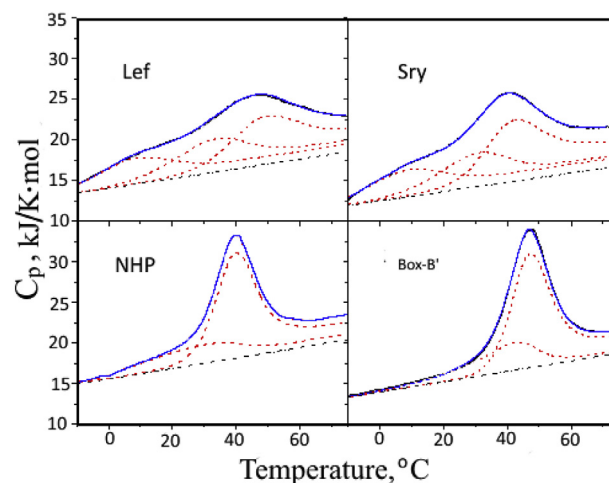

Fig. 13. Deconvolution of the partial molar heat capacity functions of four free HMG box DBDs under standard solvent conditions: 10 mM potassium phosphate, pH6.0, 100 mM KCl. The sequence-specific (SS) HMG boxes Lef and Sry melt in a series of transitions such that at physiological temperatures of 37 °C they are very significantly unfolded. The Cp/T profiles of the non-sequence-specific (NSS) HMG boxes NHP (from yeast NHP6a) and Box-B' (from hamster HMGB1) more closely exhibit a single cooperative melt. The black dotted line represents the heat capacity function of BPTI corrected to the molecular weight of the HMG box.
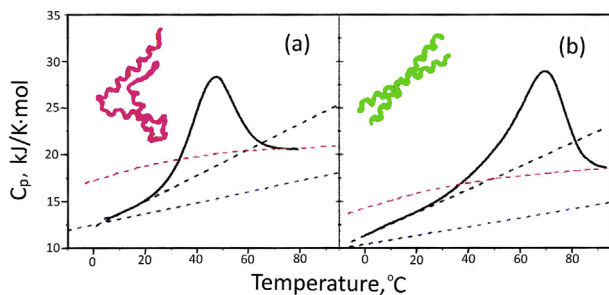

Fig. 11. The partial molar heat capacities of: (a) The HMG box from Sox5; (b) GCN4-bZIP. The initial steep rise in the heat capacities are extended through the transition with the dashed black lines. The lower black dotted lines represent the Cp/T function of the rigid BPTI protein (molecular weight corrected) and the red dotted lines represent the heat capacity of the unfolded states.
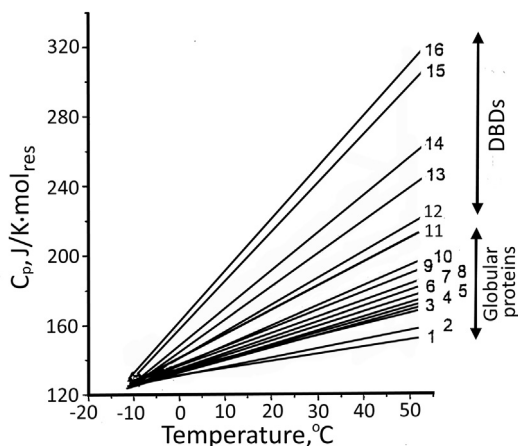
Fig. 12. The temperature dependence of the initial partial specific heat capacities of various proteins, expressed per residue: 1—Anhydrous protein; 2—BPTI; 3—Barnase; 4—Myoglobin; 5— Lysozyme; 6—Cytochrome C; 7—Ubiquitin; 8—T4 lysozyme; 9—RNase T1; 10—RNase A; 11—Engrailed; 12—HMG Sox; 13—NHP6A; 14—HMG SRY; 15—HMG Lef-79; 16—HMG Lef-86 (Privalov, 2007).

whereas the free Lef-1 protein (in green) melts with broad and weak heat absorption centered at 40—50 °C and the target 16 bp DNA melts as a sharp peak with a low temperature tail (in red), the complex melts as a very symmetrical sharp peak at a similar temperature to the DNA, i.e. a unitary rigid complex forms and there is no indication of either component in the complex melting individually. *The DBD and its DNA binding site thus form a single and stable cooperative unit in the complex.*

9. The energetic basis of forming DBD-DNA complexes

Titration of a DBD onto its target DNA in an ITC experiment yields the heat, i.e. the enthalpy (ΔH) of the interaction and, frequently, the association constant, K_a. The linear temperature dependence of ΔH represents the accompanying heat capacity change, ΔCp. Accurate knowledge of both these quantities is important, as together with the Gibbs free energy change, ΔG, (obtained as $\Delta G = -RT \cdot \ln[K_a]$) they allow derivation of a full thermodynamic profile of the binding process over a significant temperature range. However, measurement of heats of binding can be complicated by conformational changes in one or both the components. The example in Fig. 14, Panel (b) of the Lef-1 HMG box, shows ITC-measured heats of binding (broken line) that are *not* linear with temperature because with temperature rise the unbound free protein becomes increasingly unfolded but refolds on binding DNA, liberating considerable heat. However, this refolding heat at a given temperature can be estimated from DSC measurements of the free protein and this permits correction of the ITC-measured heats of binding. When these heats are subtracted from the ITC-measured heats of binding, the temperature dependence of the corrected binding enthalpy becomes linear, giving the solid line in Fig. 14, Panel (b). The slope of this corrected line represents the heat capacity change on binding fully folded protein to fully folded DNA.

Refolding of DBDs on binding their target DNA is a very general phenomenon (Spolar and Record, 1994). The very large family of bZIP DBDs contain a basic region — representing the binding element — that is almost completely unfolded in the absence of the

Fig. 14. (a) The heat capacity functions of the HMG box from LEF-1 (green), its target DNA duplex (red) and their complex (black). The protein starts to unfold from very low temperatures but on association with DNA it refolds and forms a stable complex that dissociates and unfolds cooperatively at 62 °C. (b) The observed enthalpy of association of the HMG box from LEF-1 with its target DNA measured by isothermal titration calorimetry (ITC, broken line), and the function corrected for heats of protein refolding upon bi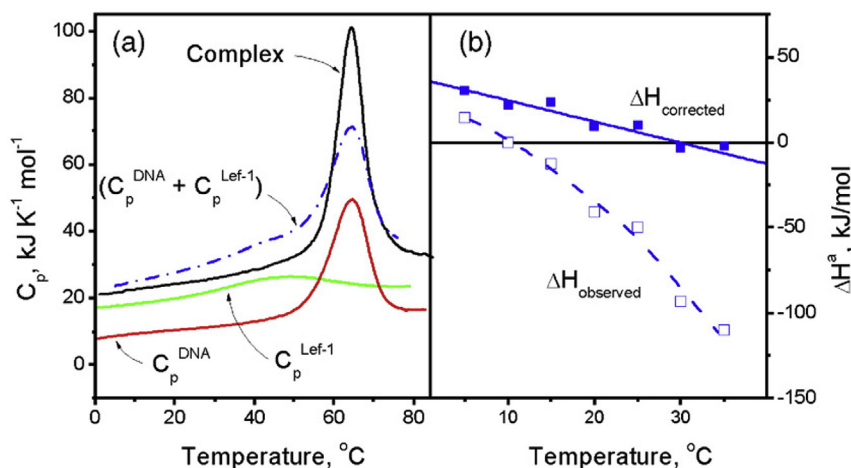nding (solid line). The corrected function corresponds to the enthalpy of association of the fully folded DBD with the DNA. (From Privalov et al., 2007).

target DNA but becomes fully α−helical when bound into the major groove (Berger et al., 1996; Dragan et al., 2004a). Even for homeo-domains, often regarded as fully folded DBDs, the actual DNA recognition element (Helix 3) is often partially unfolded in the free protein, becoming fully helical in the complex (Carra and Privalov, 1997; Dragan et al., 2006). Other cases include MYB family DBDs (Sarai et al., 1993) and zinc fingers (Hyre and Klevit, 1998; Liggins and Privalov, 2000).

10. DNA bending

A frequent feature of minor groove binding is the induction of sharp DNA bends. Bend angles can be determined using the circular permutation assay (Kim et al., 1989) and also by crystallography, although the latter can be subject to uncertainty as a result of crystal packing forces enhancing or reducing the DNA bend (Masse et al., 2002; Murphy et al., 1999). Solution methods of measuring bend angles have obvious advantages, in particular NMR studies of protein-DNA complexes (Love et al., 1995; Dow et al., 2000; Masse et al., 2002) and in this context fluorescence resonance energy transfer (FRET) is a valuable tool.

FRET measures the fluorescence energy transfer between acceptor and donor fluorophores placed on opposite ends of a target DNA duplex and its magnitude provides information on the distance between the fluorophores, i.e. the distance, R_{da}, between the ends of the duplex (Fig. 15). The DNA must be short enough to generate a FRET change of measurable magnitude but long enough to ensure no direct contact between the DBD and the fluorophores. Increases in the FRET on protein association with DNA thus char-acterize the protein-induced bending of the DNA. For a given type of DBD the observed R_{da} values are best calibrated empirically with complexes having bend angles known from structure de-terminations. A great advantage of this method is that it can be used under varying conditions of ionic strength and temperature. Furthermore, having a fluorophore attached to the DNA allows determination of DBD/DNA binding curves from changes in the anisotropy of the emission, which is particularly valuable if the affinity is high, for example in the low nM range.

11. DBD binding to the minor and major grooves of DNA

A striking feature of DNA/protein complexes is that sequence specific binding to the minor groove usually takes place at AT-rich sequences and can result in considerable DNA bending, by even more than 90°, in contrast to major groove binding (Fig. 16a). However, despite large differences in the DNA deformations caused by DBD binding to the two grooves, the Gibbs energies of binding to the minor and major grooves are fairly similar, around −40 kJ/mol

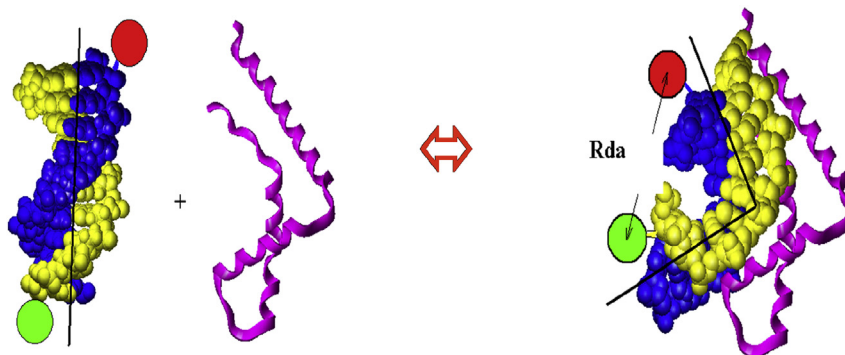**Fig. 15.** Determination of DBD induced DNA bending by FRET. Red and green circles represent donor and acceptor fluorophores attached to the 5′ ends of the two DNA strands (yellow and blue). The 16 bp duplex is shown interacting with an HMG box (magenta). The induced bending results in a reduction of the distance R_{da} i.e. an increased FRET effect (for details see Dragan and Privalov, 2008).
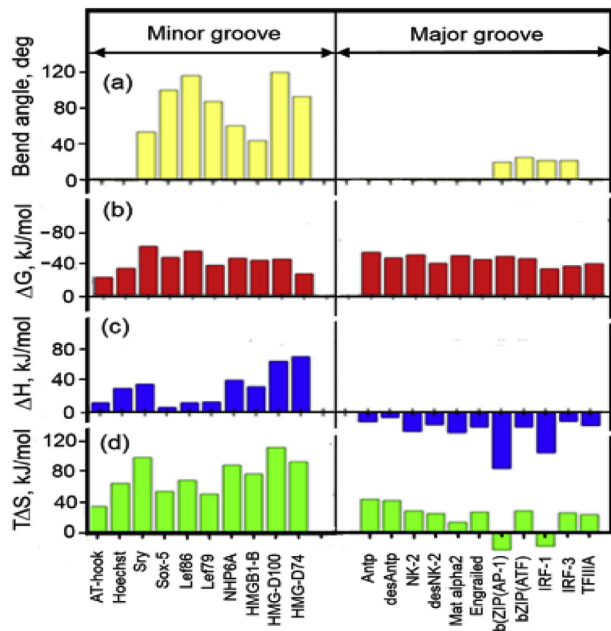
Fig. 16. Interaction of the DBDs of various transcription factors with their target DNA sequences at 20 °C in 10 mM potassium phosphate (pH 6.0), 100 mM KCl: (a) DNA bend angles induced; (b) the Gibbs energy of binding; (c) the enthalpy of binding; (d), the entropy factor of binding (for details see Privalov et al., 2007).

(K_d ~50 nM), providing stable enough DNA/DBD complexes at modest concentrations of the transcription factors (Fig. 16b). Surprisingly, the enthalpies of binding to the minor and major grooves differ qualitatively: they are positive for binding to the minor groove and negative for binding to the major groove (Fig. 16c). For binding to the major groove negative enthalpies have frequently been noted (Ladbury et al., 1994; Hyre and Spicer, 1995; Merabet and Ackers, 1995) and positive enthalpies have previously been reported for minor groove binding (O'Brien et al., 1998; Haq et al., 1997). It follows that these large differences in the enthalpies are balanced by entropy factor differences (Fig. 16d).

A negative enthalpy promotes binding, while a positive enthalpy opposes. Binding to the minor groove is therefore driven by the entropy, which is large and positive, in contrast to the entropy of binding to the major groove, which while also positive is small in magnitude. It follows that *binding of DBDs to the minor groove is an entropy driven process, while binding to the major groove is largely enthalpy driven*. It was initially suggested that the large entropy increase on minor groove binding is conformational (Privalov et al., 1999; Jen-Jakobson et al., 2000), a consequence of the resulting bent complex being a very flexible structure of elevated entropy. But this explanation is not borne out by the structures determined for such complexes (by both NMR and X-ray) that appear to have normal rigidity. Furthermore, the calorimetric observations in Fig. 14 (a) show that the Lef-1/DNA complex comprises a single rigid cooperative unit. The large entropy of minor groove binding can only result from displacement of the ordered water bound to AT-rich target sequences (see Fig. 9). The corresponding increase in the enthalpy from melting this 'ice-like' water is apparent from its net positive magnitude, although this is reduced by the negative enthalpy contribution resulting from formation of close contacts at the DNA/DBD interface in sequence-specific complexes. The net result is that the positive entropy of water displacement is more evident than the positive enthalpy in minor groove binding.

12. Electrostatic and non-electrostatic components of the DNA/DBD interaction

As DNA is a highly charged macromolecule, electrostatic interactions are involved in the formation of DNA-DBD complexes. One approach, initially proposed by Manning (1969, 1978) and implemented into the study of protein-DNA complexes by Record and colleagues (1976, 1978; 1991; Lohman et al., 1980; Lohman and Macotti, 1992), assumes that the electrostatic component of the binding energy results solely from the cratic entropy of mixing the displaced DNA counter-ions with ions in the bulk solution. According to this counter-ion condensation concept, the electrostatic component of the binding can be determined directly from the salt dependence of the association constant, K^a, as described by the linear free energy equation:

$$\log(K^a) = \log(K^a_{nel}) - N \cdot \log[\text{Salt}] \quad (13)$$

where the first term accounts for the non-electrostatic interactions and the second represents the salt dependent electrostatic interactions in which N is the total number of counter-ions released from the DNA on forming the complex. N can be written as the product $Z \cdot \psi$, where Z is the number of DNA phosphate groups that interact with the protein/peptide and ψ is the average number of cations associated with a phosphate group that are displaced on complex formation.

Since at 1 M salt concentration the second term drops to zero, analyzing the salt dependence of the association constant leads to an extrapolated value for the non-electrostatic component of binding: $\Delta G_{nel} = -2.3RT\log(K^a_{nel})$. The electrostatic component is then given as the difference from the total Gibbs energy, ΔG^a, at the solution conditions of interest: i.e. $\Delta G_{el} = \Delta G^a - \Delta G_{nel}$. As the enthalpy of electrostatic interactions is zero (Dragan et al., 2004b), ΔG_{el} is equivalent to $-T\Delta S_{el}$. The non-electrostatic entropy factor is then obtained by subtracting this from the total entropy factor: i.e $T\Delta S_{nel} = T\Delta S^a - T\Delta S_{el}$. This approach fully defines the electrostatic and non-electrostatic components of the binding free energy. Moreover, from the slope of the logarithmic dependence, $Z \cdot \psi$, one can obtain the total number of counter-ions released on forming a DBD-DNA complex and thereby estimate the number of contacts between the partners forming the complex by using the experimentally derived value of $\psi = 0.70$. This use of 'salt plots', although criticized by theoreticians, is still the only practical approach for separating the contributions of electrostatic and non-electrostatic interactions in the formation of DBD-DNA complexes. The topic is discussed in more detail in Privalov et al. (2011).

13. Specificity of DNA recognition by DBDs

The ability to separate the electrostatic and non-electrostatic components of the binding free energy and measure the induced bend angle permit a more detailed analysis of the energetic basis of the recognition process. This is illustrated by the DNA binding of sequence-specific (SS) and non-sequence-specific (NSS) HMG boxes to their optimal target and sub-optimal sequences, Fig. 17.

One can see that for the SS HMG boxes (upper panel) the electrostatic component (blue) is independent of the target sequence for a given HMG box, i.e. the phosphate-K/R interactions remain the same despite significant changes in the overall affinity and in the induced bend angle. In contrast, the non-electrostatic component of the Gibbs energy (yellow) is largest for association of a given SS DBD with its optimal target DNA and less for binding to sub-optimal sequences. In the case of NSS HMG boxes, not only the
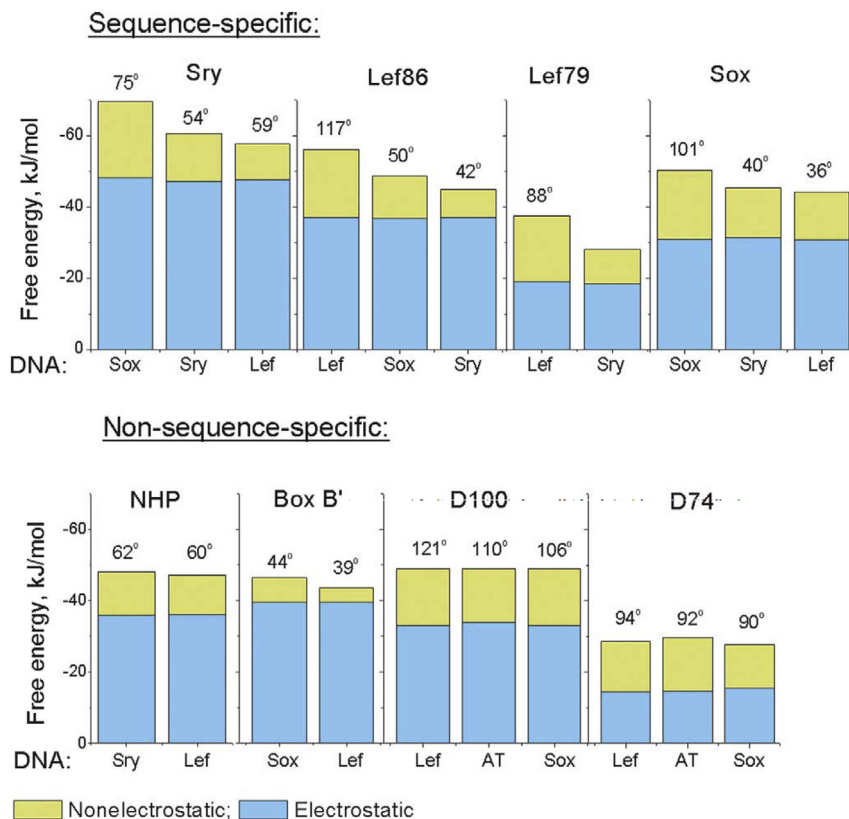
Fig. 17. The electrostatic (light blue) and non-electrostatic (yellow) components of the total Gibbs free energy of binding sequence-specific (SS) and non-sequence-specific (NSS) HMG boxes to various DNAs. For SS boxes the first DNA (on the LHS) can be considered the optimal target. The numbers above the bars indicate the induced DNA bend angles measured by FRET in the standard buffer: 10 mM potassium phosphate, pH 6.0, 100 mM KCl (for details see Dragan et al., 2004b; Privalov et al., 2009).

electrostatic component but also the non-electrostatic Gibbs energy of binding remain the same for binding a particular DBD to all the target sites: for this reason they are termed 'non-sequence-specific'. It follows therefore that *specificity of binding is determined by the non-electrostatic part of the Gibbs energy: this component results mostly from van-der-Waals interactions and H-bonds.*

The distinction between SS and NSS binding of HMG boxes lies in the non-electrostatic component and Fig. 17 shows that for SS boxes the induced bend angles are roughly in proportion to the non-electrostatic component of the Gibbs energy, i.e. bend angles are in proportion to the number of newly formed van der Waals contacts. This distinction can be visualized by comparing the packing densities of SS and NSS HMG box-DNA complexes. Fig. 18 shows that the SS boxes from Lef-1 and Sry form a large number of close contacts (in red) throughout the protein-DNA interface when bound to their target sequences, DNA^{Lef} and DNA^{Sry}, whereas such contacts are largely absent in the complexes of the NSS boxes from HMG-D74 and NHP6A.

The formation of tight DNA complexes by the SS HMG boxes, which have a very loose structure in free solution (evident from Fig. 13), suggests that the DBD should initially be flexible enough to envelop the DNA so as to form multiple contacts, i.e. *the flexibility of the DBD structure is required for its proper adjustment to the DNA structure.*

14. The contribution of DBD 'extensions'

A frequent feature of DBDs is segments (sub-domains) totally disordered in the free DBD which bind to the DNA independently of the principal folded interacting domain. These are frequently described as 'tails', 'arms' or 'extensions'. *Drosophila* Antennapedia in Fig. 10C is an example. As the tails usually include basic residues they are expected to increase the affinity of the DBD for its target DNA and in certain cases they can also affect the specificity of binding. The specificity and affinity of binding the disordered tails of DBDs to DNA can be characterized by plots of Eqn. (13) ('salt plots'). Fig. 19 gives plots for two different DBDs, one minor groove binding the other major groove binding, both of which have positively charged tails that are unstructured in free solution but bound to DNA in their complexes.

The left hand panel of Fig. 19 compares Lef79 (the fully folded HMG box domain from mouse LEF-1 protein) with Lef86 (see Fig. 10B) that additionally contains a very basic C-terminal tail of 8 residues (of which 7 are basic) that straddles the major groove on the inside of the bent DNA (Love et al. 1995). The tail is responsible for generating about 30% of the induced bend angle in the complex (Lnenicek-Allen et al. 1996). When binding to the optimal target sequence, DNA^{Lef}, the plots for the two proteins converge at 1 M NaCl where the electrostatic effects are lost and both are bound only by non-electrostatic forces, of equal magnitude for both. The Lef86 tail therefore binds only by electrostatic forces. Comparing their binding to DNA^{Lef} in 100 mM salt (log[Salt] = −1.0) shows that the tail increases the affinity by three orders in K^a, i.e. contributes ~30% of the total Gibbs energy of binding. The greater slope of the Lef86 plot reflects the increased number of electrostatic contacts in its basic tail: whereas the globular Lef79 makes 5 ionic contacts, Lef86 makes 10, i.e. tail contacts average 5 in number. As the tail contributes 18 kJ/mol to the total Gibbs energy of binding, this represents 3.6 kJ/mol per ionic link. When binding to the sub-optimal sequence, DNA^{Sry}, their

Sequence-specific

Lef86/DNA^{Lef}   Sry/DNA^{Sry}

Non-sequence-specific

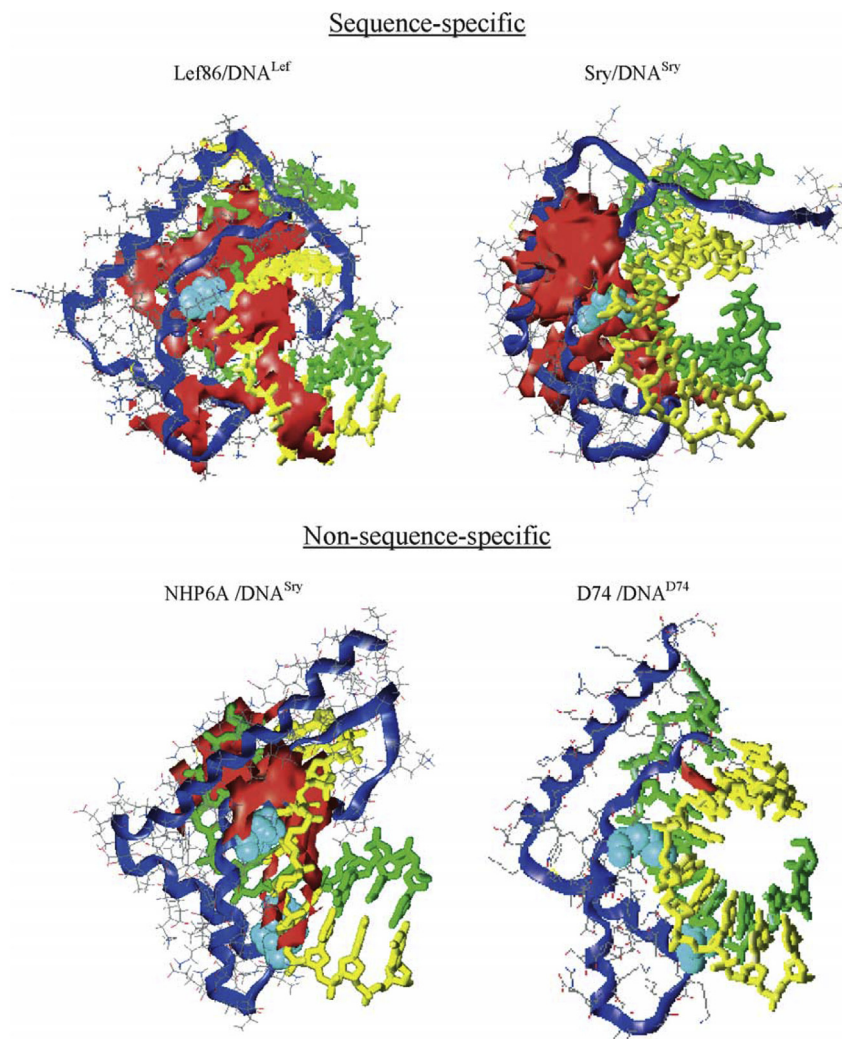NHP6A /DNA^{Sry}   D74 /DNA^{D74}

Fig. 18. Packing density displays of DNA complexes formed by two SS HMG boxes (Lef86 and Sry) and two NSS HMG boxes (yeast NHP6A and *Drosophila* D74) having similarly bent DNA. The packing density is defined as the ratio between the van der Waals envelope of a molecular volume and the total volume occupied (Richards, 1985). Red clusters are regions with a packing density >0.68. Intercalating residues in pale blue. The SS complexes form a much tighter (more complementary) interface than the NSS complexes. Packing densities were analyzed using the program MOLE (Molecular Graphics and Computation, Applied Thermodynamics, LLC) (Reproduced from Dragan et al 2004b).

slopes are the same as with the optimal target, i.e. the electrostatic contacts remain unchanged. The plots again converge at 1 M NaCl but the non-electrostatic affinity is reduced by one order in K^a due to the less intimate (complementary) Lef79/DNA^{Sry} interface than formed with the optimal DNA^{Lef} target.

The right hand panel of Fig. 19 shows a different type of basic tail: the 8-residue N-terminal extension of the *Drosophila* Antennapedia (homeodomain) DBD includes 4 basic residues and binds in the minor groove, while the recognition helix binds in the major groove (Otting et al., 1990) — see Fig. 10c. In this case the two salt plots exhibit the same slope, i.e. both Antp and desAntp are bound by the same number of ionic contacts (6 in number) and desAntp has a lower affinity throughout. It follows that the N-terminal extension of Antp is bound non-electrostatically and provides ~12% of the total free energy of Antp binding in 100 mM NaCl (Crane-Robinson et al., 2006). This conclusion accords with the structure of the complex that shows the basic residues of the extension bound deep in the minor groove, not contacting phosphate groups, and providing some specificity of recognition. It follows that the Lef-1 tail represents *an ionic non-sequence-specific entropic binding force,* whereas the Antennapedia tail uses *sequence-specific enthalpic* contacts.

15. The components of the DBD-DNA interaction

The enthalpies of binding DBDs to their target DNAs are entirely non-electrostatic in origin and this has been verified by noting that ITC-measured heats of binding HMG boxes to optimal recognition sequences are independent of the salt concentration, despite the overall changes in affinity (Dragan et al., 2004b). Negative enthalpies derive from the formation of van der Waals contacts and hydrogen bonds, while positive contributions come from the release of strongly bound water from the DNA and protein. In contrast, the entropies of binding have both an electrostatic component derived from the release of counter-ions from the DNA and a non-electrostatic component derived from conformational changes and, more significantly, the release of strongly bound water into the bulk solution.

Fig. 20 plots the component enthalpies and the electrostatic and non-electrostatic entropies of DBD binding to both the major and the minor grooves (Privalov et al., 2011). One firstly notes that the electrostatic contributions to binding ($T\Delta S_{el}$, in blue) are fairly constant throughout, in both the major and minor grooves, showing that a similar number of salt links are made for all these globular DBDs, an electrostatic component typically contributing
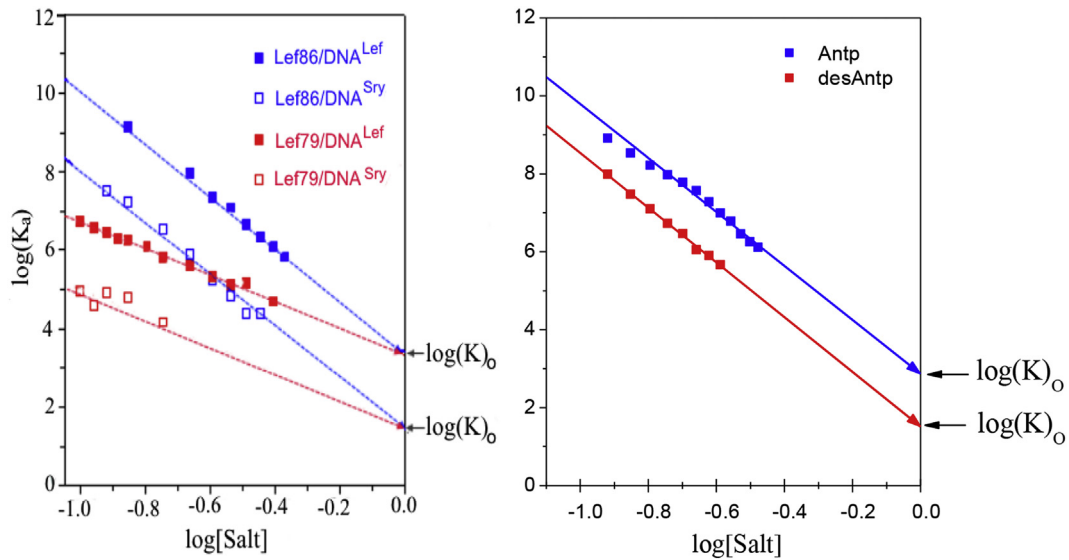
Fig. 19. The salt dependence of binding constants, K^a, (Privalov et al., 2011). (A) Dependence of log(K^a) on log [NaCl] for the HMG box from mLEF-1 (Lef86) and its truncated form lacking the 8 residue basic C-terminal tail (Lef79) binding to DNA^{Lef} (−TTCAAA−), the optimal target and also to DNA^{Sry} (−CACAAA−), a sub-optimal target. (B) Dependence of log(K^a) on log [NaCl] for the binding of the full (Antp) homeodomain from the *Drosophila* Antennapedia protein and its truncated (desAntp) form lacking the 8 residue N-terminal tail, both with the same (optimal) DNA target sequence at 20 °C.
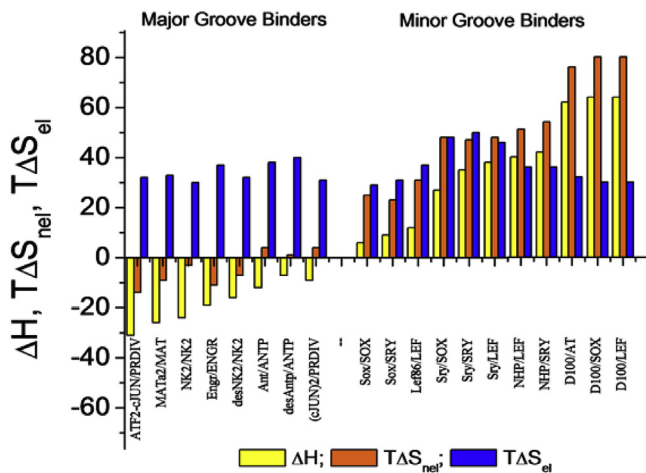

Fig. 20. Enthalpies (ΔH) and entropy factors (TΔS: nel − non-electrostatic, el −electrostatic) of binding proteins to the minor and major groove of their optimal and sub-optimal DNA target sequences at 20 °C in 10 mM potassium phosphate, pH 6.0, 100 mM KCl. SS = sequence-specific, NSS = non-sequence-specific DNA binding domains (Privalov et al., 2011).

60−70% of the total Gibbs energy. *It is a major component driving formation of DNA-protein complexes but it is a non-sequence-specific binding force.* Secondly, a rough correlation is seen between the enthalpy (ΔH, non-electrostatic, yellow) and the non-electrostatic entropy factor (TΔS$_{nel}$, orange). *For DBD binding to the major groove both these components are generally negative and the enthalpy substantially exceeds the entropy factor.* This rather small negative entropy factor results mainly from the decrease in conformational and translational freedom of the DBDs and the DNA on association, while the enthalpy represents the contribution of interfacial contacts. The net magnitude of these two factors adds to the electrostatic component to give the overall Gibbs energy of binding to the major groove.

In the case of DBD binding to the minor groove the situation is drastically different: both the enthalpy and non-electrostatic entropy factor are positive and the entropy factor dominates the enthalpy. This immediately raises the question: from where do these large positive non-electrostatic entropies and enthalpies come when DBDs bind to the minor groove of DNA, particularly the non-sequence-specific (NSS) HMG boxes such as NHP and D100? It is notable that DBD binding to the minor groove of DNA is preferentially at AT-rich sequences, i.e. to a minor groove containing ordered water. The only possible explanation for these large positive enthalpies and entropies is that *removal of this ordered minor groove water upon protein binding gives rise to both a large positive enthalpy and an especially large positive entropy of protein binding,* i.e. dehydration of the DNA is critical for protein binding to the minor groove. If this tightly bound ordered water has the thermodynamic properties characteristic of ice (as discussed in Section 7, above) then its removal, i.e. the 'melting' of this ice, should essentially be a zero free energy process (ΔG~0), as the temperature is close to 273 K (Grunwald and Steel, 1995). Thus the enthalpy required should equate to the accompanying increase in the non-electrostatic entropy factor, TΔS$_{nel}$. This is approximately the case for the NSS HMG boxes (the last 5 examples on the RH side of Fig. 20). However, for the SS HMG boxes such as Sox and Lef the non-electrostatic entropy is more positive than the enthalpy: this is the result of a negative enthalpy contribution from the formation of specific van der Waals contacts/H-bonds at the protein/DNA interface.

16. Non-sequence-specific DNA binding

Non-sequence-specific HMG boxes, such as *Drosophila* HMG-D and yeast NHP6A, bind to the minor groove and their Gibbs energies are dominated by the electrostatic contribution, to which is added a non-electrostatic component in which the entropy increase from released water dominates a positive enthalpy (see Fig. 20). Furthermore, the heat capacity changes are significantly negative, implying formation of a hydrophobic interface (ΔCp = −1.0 kJ/K/mol for HMG-D74). The binding of TBP (TATA box protein) to its AT-rich minor groove target is likewise dominated by a large entropy increase over a positive enthalpy (O'Brien et al., 1998) and the same is true for the DNA packaging protein Sso7d from the hyperthermophilic archaeon *S. solfataricus* (Lundback

et al., 1998). For both TBP and Sso7d, ΔCp is strongly negative (-2.1 and $-1.0\,kJ/K/mol$, respectively). The above proteins are sometimes designated 'architectural' transcription factors as their role is to bind DNA tightly by forming a hydrophobic interface without any sequence preference: distortion of the DNA shape follows. This type of non-sequence-specific binding is characterized by the formation of a substantial hydrophobic interface but without base-specific contacts. Such interactions give rise to substantially negative ΔCp values, which come largely from protein to sugar ring contacts.

This type of non-sequence-specific DBD-DNA interaction can be contrasted with cases in which a sequence-specific DBD binds to a sequence very different from its optimal target: an example being the bacterial Cro repressor. When Cro binds *specifically* to OR3 — a natural target — by inserting a recognition helix into the major groove, binding is driven by a large entropy increase (Takeda et al., 1992) and the thermodynamic binding signature is similar to desAntp (Fig. 15), as expected for a similar recognition mechanism. However, when Cro binds to *non-specific* DNA, the enthalpy is more positive and *not* temperature dependent (i.e. $\Delta Cp\sim0$), meaning that no specific hydrophobic interface is formed. Cro is then held entirely by electrostatic forces and these give rise to the substantial entropy that drives its non-specific binding (Takeda et al., 1992). Similarly, it has been shown that non-specific binding of the *trp* repressor is characterized by a zero heat capacity change, $\Delta Cp\sim0$, (Ladbury et al., 1994). A structurally defined non-specific binding mode has likewise been characterized for the dimeric *lac* repressor headpiece in which the amino acids responsible for sequence recognition switch their binding to contact only phosphates, i.e. all sequence recognition is lost (Kalodimos et al., 2004). In the case of the HOXD9 protein binding to non-cognate DNA, intermolecular paramagnetic relaxation enhancement (PRE) measurements show the presence of transient intermediates *lacking* close contacts with the DNA but having structures similar to that of the specific complex (Iwahara and Clore, 2006). Such examples of non-specific binding of DBDs correspond to the established concept that non-specific binding to DNA is largely electrostatic and the DBD is disengaged from intimate contact with the target sequence, i.e. van der Waals contacts and H-bonds are lost. This is the favored model for the sliding of DBDs along DNA in a search mode, for example the restriction enzyme BamH1 (Viadiu and Aggarwal, 2000).

17. Rigidity of the DNA double helix

DNA free in solution has been characterized by the worm-like chain (WLC) model of Kratky and Porod as an elastic rod with a persistence length, L_p, of ~50 nm (~150 bp) (Taylor and Hageman, 1990; Baumann et al., 1997; Yuan et al., 2008). The free DNA double helix is therefore a rather rigid rod, the stiffness of which is postulated to result from favorable base pair stacking interactions that contract the duplex, opposed by phosphate charge repulsions that expand it (Peters and Maher, 2010). It should therefore be difficult to bend DNA by binding DBDs and the worm-like chain model predicts a free energy expenditure of about 70 kJ/mol in bending a 10 bp duplex through 50°, that is, about 1.5 kJ per degree of bend (Landau and Lifshitz, 1970).

Two main mechanisms of protein-induced DNA bending have been considered: the asymmetric neutralization of DNA phosphates and the insertion of protein side chains between DNA bases to generate kinks (see for review Privalov et al., 2009). The idea that asymmetric shielding of DNA phosphates by the positive charges of bound protein might be a driving force in bending DNA was originally advanced by Mirzabekov and Rich in 1979. The most obvious example of bending by such a mechanism is the DNA in a nucleosome, which forms a continuous superhelix of ~80 bp/turn without kinking (Luger et al., 1997). A further example is provided by the

unfolded N- and C-terminal tails of the DBDs of certain HMG box proteins (such as Lef-1, see Figs. 10 and 19, NHP6A and HMG-D, Dragan et al., 2003) that carry multiple positive charges and lie across the major groove to induce considerable bending (Love et al., 1995; Lnenicek-Allen et al. 1996; Murphy et al., 1999; Masse et al., 2002; Dragan et al. 2003).

Sharp local bends in DNA are normally generated by kinking, i.e. introduction of a large roll without loss of base pairing (first proposed by Crick and Klug, 1975, the concept was developed by Cloutier and Widom, 2004). This is the mechanism adopted by the saddle-shaped TATA box binding protein (TBP) molecule (Kim et al., 1993) and by the β–loop arms of the integration host factor (IHF) homodimer (Rice et al., 1996), as well as by the HMG box proteins. Such proteins typically bind in the minor groove at AT-rich sites, forcing the DNA to bend away from the protein toward the major groove (Love et al., 1995; Werner et al., 1995). Analysis of HMG box binding to DNA therefore presents an opportunity to investigate the energetic consequences of bending (more precisely, kinking) DNA through sharp angles.

The data in Figs. 15 and 17 show HMG box binding to be characterized by an unfavorably positive enthalpy but a very favorable positive entropy and it is now clear that this results from removal of ordered water from the AT rich minor groove. How then do these processes affect the magnitude of the bend angle generated? A good example is the DBD from mLEF-1 (Lef86), the salt plots for which (Fig. 19, left hand panel) exhibit the same slope for binding to the optimal recognition sequence (DNA^{Lef}) and to a sub-optimal sequence (DNA^{Sry}), meaning that the same number of ionic contacts are made with both DNA sequences. However, the bend angles observed are very different: 117° and 39°, respectively. It follows that the bend angle is independent of the electrostatic component of the binding energy, i.e. only the non-electrostatic components of the binding reaction are important for determining the bend angle. Fig. 21 therefore plots the enthalpy and the non-electrostatic entropy factors, together with the resulting Gibbs energies, against the observed bend angle for sequence specific HMG boxes (SS HMG DBDs, left hand panel) and non-sequence-specific HMG boxes (NSS HMG DBDs, right hand panel) binding to optimal and sub-optimal DNA sequences.

For SS HMG box DBDs the optimal target sequence, i.e. that having the highest affinity, always generates the greatest bend angle. Such complexes involve close contacts between the DBD and the DNA resulting in total exclusion of minor groove water and thus a large positive contribution to the entropy and enthalpy. Fig. 21 (LH panel) shows that their non-electrostatic Gibbs energies (ΔG_{nel}) become slightly more *negative* as the bend angle increases, i.e. the affinity rises slightly as the bend angle increases. This situation is the result of the binding surface of the SS HMG boxes having precise complementarity with that of the widened minor groove of the bent DNA, particularly with the optimal target sequence.

With the NSS HMG box DBDs, for which base-specific van der Waals contacts with the DNA are not made (Murphy et al., 1999), as the bend increases there is a substantial increase in the positive enthalpy of binding and also in the positive non-electrostatic entropy. These effects are due to the gradual exclusion of water from the interface as a greater bend angle is forced in — and there is no compensating negative enthalpy from interfacial base contacts, as occurs with the SS boxes. The net result is nevertheless that ΔG_{nel} becomes slightly more *negative* as the bend angle increases.

Overall, it is clear that sharp bending of DNA by wedge insertion from the minor groove side does *not* require increasing amounts of free energy to generate larger bend angles. This was initially a surprising result bearing in mind the rigidity of naked B-form DNA. It follows that the minor groove binding of HMG boxes must result
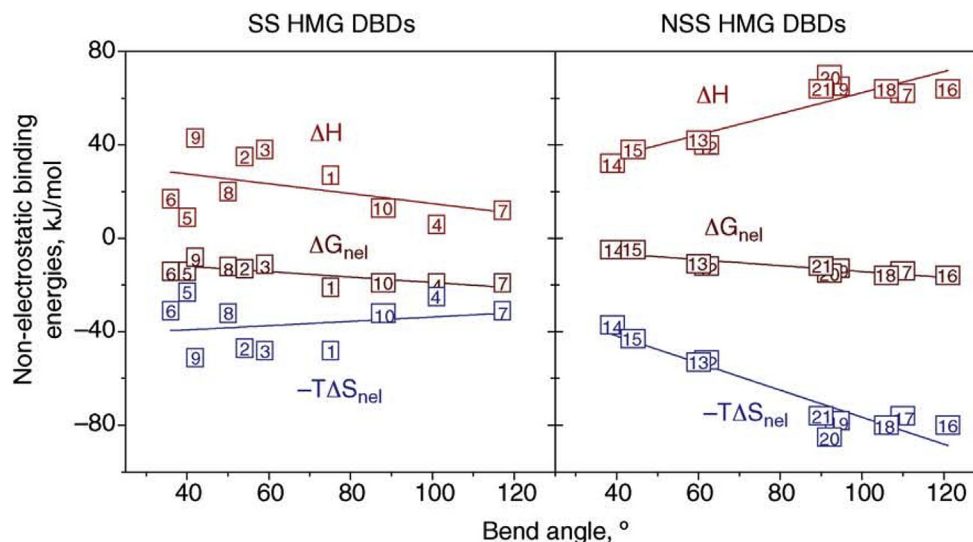
Fig. 21. The non-electrostatic components of the binding free energy of HMG boxes plotted against the induced DNA bend angles. For sequence-specific (SS) boxes the optimal target, DNA[OP], or sub-optimal targets, DNA[SO], are so designated. For non-sequence specific (NSS) boxes the different target sequences are simply designated, DNA, because the bend angles are independent of the sequence (see Fig. 17). **SS DBDs: 1** Sry/DNA[OP], **2** Sry/DNA[SO], **3** Sry/DNA[SO], **4** Sox/DNA[OP], **5** Sox/DNA[SO], **6** Sox/DNA[SO], **7** Lef86/DNA[OP], **8** Lef86/DNA[SO], **9** Lef86/DNA[SO], **10** Lef79/DNA[OP]. **NSS DBDs: 12** NHP6A/DNA, **13** NHP6A/DNA, **14** Box-B'/DNA, **15** Box-B'/DNA, **16** D100/DNA, **17** D100/DNA, **18** D100/DNA, **19** D74/DNA, **20** D74/DNA, **21** D74/DNA. For more details see Privalov et al. 2009.

in loss of stiffness in the DNA: but how does this come about? There is loss of a single base pair stacking interaction, though this must be partially compensated by contacts to the inserted protein sidechain, and although the electrostatic stiffening might be weakened by the widened minor groove, it would be increased from the compressed major groove.

However, another possible explanation of the loss of stiffness in the duplex presents itself: the ordered water in the minor groove — illustrated in Fig. 9b — plays a major role in generating the rigidity of free DNA and its displacement by HMG boxes renders the duplex intrinsically more flexible. Minor groove bound water has previously been implicated in contributing to DNA rigidity, based on the observation of anisotropy in the bending: a shorter persistence length for bending towards the major groove than towards the minor groove. This difference was attributed to a greater ease of extruding water from the major groove, relative to the minor groove (Ma and van der Vaart, 2016). The rigidity of B-form DNA in free solution thus appears to depend principally on minor groove hydration: *its loss on protein binding is not costly in free energy terms and results in loss of stiffness in the DNA, permitting easy kinking.*

18. Conclusions

1. Despite the widely held belief to the contrary, the enthalpic and entropic contributions of the AT base pair to maintenance of the DNA double helix structure significantly exceed those of the CG base pair and both are temperature dependent.
2. The temperature dependence of the enthalpic contributions of the AT and CG base pairs comes from the heat capacity increment on duplex dissociation, identical for both base pairs as it results from exposure of the very similar apolar bases to water upon DNA duplex dissociation.
3. The greater stabilizing effect of the CG base pair results not from its larger enthalpic contribution but from its smaller entropic contribution — in comparison with that of the AT base pair. The larger enthalpy and entropy contributions of the AT base pair result from the water fixed by its polar groups in the minor groove of the double helix that is lost on strand dissociation.

4. The DNA stabilizing effect of water fixed by the minor groove depends on the precise arrangement of the AT base pairs, resulting in sequence dependence of the thermodynamic parameters of AT pairs, a phenomenon described by the 'nearest neighbour (NN) interactions' model.
5. The dependence of DNA thermostability on the size and concentration of the duplex is a consequence of the translation entropy resulting from appearance of a new kinetic unit on duplex dissociation and is expressed by the stoichiometric correction term, $R\ln\{2/[N_o]\}$.
6. The DNA-binding-domains, DBDs, of many transcription factors are incompletely folded at physiological temperatures but refold completely on binding to their cognate target DNA sequences, generating a single stable cooperative domain.
7. Binding of DBDs to the major and minor grooves is characterized by similar Gibbs energies but by very different enthalpies and entropies. Whilst binding to the major groove is largely enthalpy driven, minor groove binding is entropy driven as a result of releasing the ordered water bound to AT pairs.
8. The electrostatic component of binding is entirely a result of the entropy increase from the release of counter-ions, i.e. the ionic bonds formed have no enthalpic component. This allows salt plots to be used to separate the electrostatic and non-electrostatic contributions to the Gibbs energies of binding DBDs to DNA.
9. The rigidity of B-form DNA results from water immobilized in its minor groove and its loss on DBD binding renders the duplex sufficiently flexible to enabling bending without free energy expenditure.

with proteins was initially funded by the Russian Academy of Sciences and then by the NIH of the USA. The authors especially note the contributions of Jamlet Monaselidze; Valerian Plotnikov; George Privlaov, Vincent Cavina and Rasty Russell in developing the Micro- and Nano- Calorimetric instrumentation that opened up the prospect of experimental investigations of the energetic bases of biological macromolecules. The authors are also grateful to many colleagues who over years have participated in the experimental study of DNA and its complexes: T.M. Birshtein; K.J. Breslauer; V.V. Filimonov; A.I. Dragan; I. Jelesarov; C.A.S.A. Minetti; O.B. Ptitsyn; C.M. Read; D.P. Remeta and P. Vaitiekunas.

References

Allawi, H.T., SantaLucia Jr., J., 1997. Thermodynamics and NMR of internal G: T mismatches in DNA. Biochemistry 36, 10581—10594.

Arai, S., Chatake, T., Ohhara, T., Kurihara, K., Tanaka, I., Suzuki, N., 2005. Complicated water orientations in the minor groove of the B-DNA decamer d(CCAT-TAATGG)₂. Nucleic Acids Res. 33, 3017—3024.

Avery, O., MacLeod, C., McCarty, M., 1944. Studies on the chemical nature of the substance inducing transformation of pnemococcal types. J. Exp. Med. 79, 137—157.

Baumann, C.G., Smith, S.B., Bloomfield, V.A., Bustamante, C., 1997. Ionic effects on the elasticity of single DNA molecules. Proc. Natl. Acad. Sci. USA 94, 6185—6190.

Berger, C., Jelesarov, I., Bossard, H.R., 1996. Coupled folding and site-specific binding of the GCN4-bZIP transcription factor to the AP-1 and ATF/CREB DNA sites studied by microcalorimetry. Biochemistry 35, 14984—14991.

Borer, P.N., Dengler, B., Tinoco Jr., I., Uhlenbeck, O.C., 1974. Stability of ribonucleic acid double-stranded helices. J. Mol. Biol. 86, 843—853.

Breslauer, K.J., Frank, R., Blocker, H., Marky, L.A., 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA 83, 3746—3750.

Carra, J.H., Privalov, P.L., 1997. Energetics of folding and DNA binding of the MATα2 homeo-domain. Biochemistry 36, 526—535.

Chalikian, T.V., Volker, J., Plum, G.E., Breslauer, K.J., 1999. A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques. Proc. Natl. Acad. Sci. USA 96, 7853—8758.

Chandler, D., Pratt, L.R., 1976. Statistical mechanics of chemical equilibria and intermolecular structures of nonrigid molecules in condensed phase. J. Chem. Phys. 65, 2925—2940.

Chargaff, E., 1950. Dissymmetry in nucleotide sequence of desoxypentose nucleic acids. J. Biol. Chem. 192, 223—230.

Cloutier, T.E., Widom, J., 2004. Spontaneous sharp bending of double-stranded DNA. Mol. Cell 14, 355—362.

Crane-Robinson, C., Dragan, A.I., Privalov, P.L., 2006. The extended arms of DNA-binding domains: a tale of tails. Trends Biochem. Sci. 31, 547—552.

Crick, F.H.C., Klug, A., 1975. Kinky helix. Nature 255, 530—533.

Dow, L.K., Jones, D.N.M., Wolfe, S.A., Verdine, G.L., Churchill, M.E.A., 2000. Structural studies of the high mobility group globular domain and basic tail of HMG-D bound to disulfide cross-linked DNA. Biochemistry 39, 9725—9736.

Dragan, A.I., Frank, L., Liu, Y., Makeyeva, E.N., Crane-Robinson, C., Privalov, P.L., 2004a. Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites. J. Mol. Biol. 343, 865—878.

Dragan, A.I., Read, C.M., Makeyeva, E.N., Milgotina, E.I., Churchill, M.E., Crane-Robinson, C., Privalov, P.L., 2004b. DNA binding and bending by HMG boxes: energetic determinants of specificity. J. Mol. Biol. 343, 371—393.

Dragan, A.I., Privalov, P.L., 2008. Use of fluorescence resonance energy transfer (FRET) in studying protein-induced DNA bending. Methods Enzymol. 450, 185—199.

Dragan, A.I., Li, Z., Makeyeva, E.N., Milgotina, E.I., Liu, Y., Crane-Robinson, C., Privalov, P.L., 2006. Forces driving the binding of homeodomains to DNA. Biochemistry 45, 141—151.

Dragan, A.I., Klass, J., Read, C., Churchill, M.E., Crane-Robinson, C., Privalov, P.L., 2003. DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution. J. Mol. Biol. 331, 795—813.

Drew, H.R., Dickerson, R.E., 1981. Structure of a B-DNA dodecamer. III. Geometry of hydration. J. Mol. Biol. 151, 535—556.

Egli, M., Tereshko, V., Teplova, M., Minasov, G., Joachimiak, A., et al., 1998. X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. Biopolymers 48, 234—252.

Filimonov, V.V., 1980. The thermodynamics of conformation transitions in polynucleotides. In: Hinz, H.-J. (Ed.), Thermodynamic Data for Biochemistry and Biotechnology. Springer-Verlag, Berlin, pp. 377—401.

Filimonov, V.V., Privalov, P.L., 1978. Thermodynamics of base interaction in (A)ₙ and (AU)ₙ. J. Mol. Biol. 122, 465—470.

Finkelstein, A.V., Janin, J., 1989. The price of lost freedom; entropy of bimolecular complex formation. Protein Eng. 3, 1—3.

Gamov, G.A., Ycas, M., 1955. Statistical correlation of protein and ribonucleic acid composition. Proc. Natl. Acad. Sci. USA 41, 1011—1019.

Grunwald, E., Steel, C., 1995. Solvent reorganization and thermodynamic enthalpy-entropy compensation. J. Amer. Chem. Soc. 117, 5687—5692.

Gurney, R.W., 1953. Ionic Processes in Solution. McGraw-Hill, New York.

Haq, I., Ladbury, J.E., Chowdhry, B.Z., Jenkins, T.C., Chaires, J.B., 1997. Specific binding of Hoechst 33258 to the d(CGCAAATTTGCG)₂ duplex: calorimetric and spectroscopic studies. J. Mol. Biol. 271, 244—257.

Holbrook, J.A., Capp, M.W., Saecker, R.M., Record Jr., M.T., 1999. Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. Biochemistry 38, 8409—8422.

Hyre, D.E., Klevit, R.E., 1998. A disorder-to-order transition coupled to DNA binding in the essential zinc-finger DNA-binding domain of yeast ADR1. J. Mol. Biol. 279, 929—943.

Hyre, D.E., Spicer, L.D., 1995. Thermodynamic evaluation of binding interactions in the methionine repressor system of Escherichia coli using isothermal titration calorimetry. Biochemistry 34, 3212—3221.

Iwahara, J., Clore, G.M., 2006. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. Nature 440, 1227—1230.

Janin, J., Chothia, C., 1990. The structure of protein-protein recognition sites. J. Biol. Chem. 265, 16027—16030.

Janin, J., 1995. Elusive affinities. Proteins Struct. Funct. Genet. 21, 30—39.

Jelesarov, I., Crane-Robinson, C., Privalov, P.L., 1999. Energetics of HMG box interactions with DNA: thermodynamic description of the target DNA duplexes. J. Mol. Biol. 294, 981—955.

Jen-Jacobson, L., Engler, L.E., Jacobson, L.A., 2000. Structural and thermodynamic strategies for site-specific DNA binding proteins. Structure 8, 1015—1023.

Johannesson, H., Halle, B., 1998. Minor groove hydration of DNA in solution: dependence on base composition and sequence. J. Amer. Chem. Soc. 120, 6859—6870.

Kalodimos, C.G., Biris, N., Bonvin, A.M.J.J., Levandoski, M.M., Guennuegues, M., Boelens, R., Kaptein, R., 2004. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305, 386—389.

Kauzmann, W., 1959. Some factors in the interpretation of protein denaturation. Adv. Prot. Chem. 14, 1—63.

Kim, J.L., Nikolov, D.B., Burley, S.K., 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature 365, 520—527.

Kim, J., Zwieb, C., Wu, C., Adhya, S., 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene 85, 15—23.

Kopka, M.L., Fratini, A.V., Drew, H.R., Dickerson, R.E., 1983. Ordered water structure around a B-DNA dodecamer. A quantitative study. J. Mol. Biol. 163, 129—146.

Krakauer, H., Sturtevant, J.M., 1968. Heats of the helix—coil transitions of the poly A—poly U complexes. Biopolymers 6, 491—512.

Ladbury, J.E., Wright, J.G., Sturtevant, J.M., Sigler, P.B., 1994. A thermodynamic study of the trp repressor-operator interaction. J. Mol. Biol. 238, 669—681.

Landau, L.D., Lifshitz, E.M., 1970. Theory of elasticity. In: Course of 'Theoretical Physics', second ed., vol. 7, p. 82 Pergamon, Oxford.

Liepinsh, E., Otting, G., Wuthrich, K., 1992. NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. Nucleic Acids Res. 20, 6549—6553.

Liggins, J.R., Privalov, P.L., 2000. Energetics of the specific binding interaction of the first three zinc fingers of the transcription factor TFIIIA with its cognate DNA sequence. Proteins (Suppl. 4), 50—62.

Lohman, T.H., Dehaseth, P.L., Record Jr., M.T., 1980. Pentalysine-deoxyribonucleic acid interactions: a model for the general effects of ion concentrations on the interaction of proteins with nucleic acids. Biochemistry 19, 3522—3530.

Lohman, T.H., Macotti, D.P., 1992. Thermodynamics of ligand-nucleic acid interactions. Methods Enzymol. 212, 400—424.

Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R., Wright, P.E., 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature 376, 791—795.

Lnenicek-Allen, M., Read, C.M., Crane-Robinson, C., 1996. The DNA bend angle and binding affinity of an HMG box increased by the presence of short terminal arms. Nucleic Acids Res. 24, 1047—1051.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389, 251—260.

Lundback, T., Hansson, H., Knapp, S., Ladenstein, R., Härd, T., 1998. Thermodynamic characterization of non-sequence specific DNA-binding by the Sso7d protein from Sulfolobus solfataricus. J. Mol. Biol. 276, 775—786.

Ma, N., van der Vaart, A., 2016. Anisotropy of B-DNA groove bending. J. Amer. Chem. Soc. 138, 9951—9958.

Makhatadze, G., Kim, K.-S., Woodward, C., Privalov, P.L., 1993. Thermodynamics of bpti folding. Protein Sci. 2, 2028—2036.

Makhatadze, G.I., Privalov, P.L., 1995. Energetics of protein structure. Adv. Protein Chem. 47, 307—425.

Manning, G.S., 1969. Limiting laws and counterion condensation in polyelectrolyte solutions. i. Colligative properties. J. Chem. Phys. 51, 924—933.

Manning, G.S., 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q. Rev. Biophys. 11, 179—246.

Marmur, J., Doty, P., 1962. Determination of the base composition of deoxyribonucleic acid from its thermal melting temperature. J. Mol. Biol. 5, 109—118.

Masse, J.E., Wong, B., Yen, Y.-M., Allain, F.H.T., Johnson, R.C., Feigon, J., 2002. The S. cerevisiae architectural HMGB protein NHP6A complexed with DNA; DNA and protein conformation changes upon binding. J. Mol. Biol. 323, 263—294.

Merabet, E., Ackers, G.K., 1995. Calorimetric analysis of λ cI repressor binding to DNA operator sites. Biochemistry 34, 8554–8563.

Mirzabekov, A.D., Rich, A., 1979. Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending. Proc. Natl. Acad. Sci. USA 76, 1118–1121.

Milev, S., Gorfe, A.A., Karshikoff, A., Clubb, R.T., Bosshard, H.R., Jelesarov, I., 2003. Energetics of sequence-specific protein-DNA association: conformational stability of the DNA binding domain of integrase Tn916 and its cognate DNA duplex. Biochemistry 42, 3492–3502.

Murphy IV, F.V., Sweet, R.M., Churchill, M.E.A., 1999. The structure of a chromosomal high mobility group protein – DNA complex reveals sequence-neutral mechanisms important for non-sequence specific DNA recognition. EMBO J. 18, 6610–6618.

O'Brien, R., DeDecker, B., Fleming, K.G., Sigler, P.B., Ladbury, J.E., 1998. The effects of salt on the TATA binding protein-DNA interaction from a hyperthermophilic Archaeon. J. Mol. Biol. 279, 117–125.

Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J., Wuthrich, K., 1990. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. 9, 3085–3092.

Peters, J.P., Maher III, L.J., 2010. DNA curvature and flexibility in vitro and in vivo. Q. Rev. Biophys. 43, 23–63.

Privalov, P.L., Monaselidze, D.R., Mrevlishvili, G.M., Magaldadze, V.A., 1965. Intramolecular heat of fusion of macromolecules. Sov. Phys. JTP 20, 1393–1396.

Privalov, P.L., Ptitsyn, O.B., Birshtein, T.M., 1969. Determination of stability of the DNA double helix in an aqueous medium. Biopolymers 8, 559–571.

Privalov, P.L., Gill, S.J., 1988. Stability of protein structure and hydrophobic interaction. Adv. Prot. Chem. 39, 191–234.

Privalov, P.L., Jelesarov, I., Read, C.M., Dragan, A.I., Crane-Robinson, C., 1999. The energetics of HMG box interactions with DNA: thermodynamics of the DNA binding of the HMG box from mouse Sox-5. J. Mol. Biol. 294, 997–1013.

Privalov, P.L., Tamura, A., 1999. Comments on the comments. Protein Eng. 12, 187.

Privalov, P.L., 2007. Thermodynamic problems in structural molecular biology. Pure Appl. Chem. 79, 1445–1462.

Privalov, P.L., Dragan, A.I., Crane-Robinson, C., Breslauer, K.J., Remeta, D.P., Minetti, S.A., 2007. What drives proteins into the major or minor grooves of DNA? J. Mol. Biol. 365, 1–9.

Privalov, P.L., Dragan, A.I., 2007. Microcalorimetry of biological macromolecules. Biophys. Chem. 126, 16–24.

Privalov, P.L., Dragan, A.I., Crane-Robinson, C., 2009. The cost of DNA bending. Trends Biochem. Sci. 34, 464–470.

Privalov, P.L., Dragan, A.I., Crane-Robinson, C., 2011. Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. Nucleic Acids Res. 39, 2483–2491.

Privalov, P.L., 2012. Microcalorimetry of Macromolecules. The Physical Bases of Biological Structures. Wiley & Sons.

Privalov, P.L., Crane-Robinson, C., 2018. Translational entropy and DNA duplex stability. Biophys. J. 14, 15–20.

Prive, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.S., Kopka, M.L., Dickerson, R.E., 1987. Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. Science 238, 498–504.

Record Jr., M.T., Lohman, M.L., De Haseth, P., 1976. Ion effects on ligand-nucleic acid interactions. J. Mol. Biol. 107, 145–158.

Record Jr., M.T., Anderson, C.F., Lohman, T.M., 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. Q. Rev. Biophys. 11, 103–178.

Record Jr., M.T., Ha, J.-H., Fisher, M.A., 1991. Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. Methods Enzymol. 208, 291–343.

Rice, P.A., Yang, S., Mizuuchi, K., Nash, H.A., 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 87, 1295–1306.

Richards, F.M., 1985. Calculation of molecular volumes and areas for structures of known geometry. Meth. Enzymol. 115, 440–464.

SantaLucia Jr., J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA 95, 1460–1465.

Sarai, A., Uedaira, H., Morii, H., Yasukawa, T., Ogata, K., Nishimura, Y., Ishii, S., 1993. Thermal stability of the DNA-binding domain of the Myb oncoprotein. Biochemistry 32, 7759–7764.

Searle, M.S., Williams, D.H., Gerhard, U., 1992. The cost of conformational order: entropy changes in molecular associations. J. Am. Chem. Soc. 114, 10697–10704.

Spolar, R.S., Record Jr., M.T., 1994. Coupling of local folding to site-specific binding of proteins to DNA. Science 263, 777–784.

Shui, X., Sines, C.C., McFail-Isom, L., Van Derveer, D., Williams, L.D., 1998. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. Biochemistry 37, 16877–16887.

Sugimoto, N., Nakano, S., Yoneyama, M., Honda, K., 1996. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Res. 24, 4501–4505.

Takeda, Y., Ross, P.D., Mudd, C.P., 1992. Thermodynamics of Cro protein-DNA interactions. Proc. Natl. Acad. Sci. USA 89, 8180–8184.

Tamura, A., Privalov, P.L., 1997. The entropy cost of protein association. J. Mol. Biol. 273, 1048–1060.

Taylor, W.H., Hagermen, P.J., 1990. Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. J. Mol. Biol. 212, 363–376.

Tereshko, V., Minasov, G., Egli, M., 1999. The Dickerson-Drew B-DNA dodecamer revisited at atomic resolution. J. Am. Chem. Soc. 121, 470–471.

Tidor, B., Karplus, M., 1994. The contribution of vibration entropy to molecular association. The dimerization of insulin. J. Mol. Biol. 238, 405–414.

Vaitiekunas, P., Crane-Robinson, C., Privalov, P.L., 2015. The energetic basis of the DNA double helix: a combined microcalorimetric approach. Nucleic Acids Res. 43, 8577–8589.

Viadiu, H., Aggarwal, A.K., 2000. Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. Mol. Cell. 5, 889–895.

Vesnaver, G., Breslauer, K.J., 1991. The contribution of single-stranded order to the thermodynamics of duplex formation. Proc. Natl. Acad. Sci. USA 88, 3569–3573.

Watson, J.D., Crick, F.H.C., 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737–738.

Werner, M.H., Huth, J.R., Gronenborn, A.M., Clore, G.M., 1995. Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell 81, 705–714.

Yu, Y.B., Lavigne, P., Kay, C.M., Hodges, R.S., Privalov, P.L., 1999. Contribution of translational and rotational entropy to the unfolding of a dimeric coiled-coil. J. Phys. Chem. 103, 2270–2278.

Yuan, C., Chen, H., Lou, X.W., Archer, L.A., 2008. DNA bending stiffness on small length scales. Phys. Rev. Lett. 100, 018102.