Thermodynamics of Sequence-Specific Binding of PNA to DNA†

Tommi Ratilainen,*,‡ Anders Holmén,§ Eimer Tuite,‡ Peter E. Nielsen,|| and Bengt Nordén‡

Department of Physical Chemistry, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden, Substance Analysis, Pharmaceutical R&D, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden, and Department of Biochemistry B, Center for Biomolecular Recognition, The Panum Institute, Blegdamsvej 3c, DK-2200 N Copenhagen, Denmark

Received January 7, 2000; Revised Manuscript Received April 13, 2000

ABSTRACT: For further characterization of the hybridization properties of peptide nucleic acids (PNAs), the thermodynamics of hybridization of mixed sequence PNA−DNA duplexes have been studied. We have characterized the binding of PNA to DNA in terms of binding affinity (perfectly matched duplexes) and sequence specificity of binding (singly mismatched duplexes) using mainly absorption hypochromicity melting curves and isothermal titration calorimetry. For perfectly sequence-matched duplexes of varying lengths (6−20 bp), the average free energy of binding ($\Delta G°$) was determined to be -6.5 ± 0.3 kJ mol^{-1} bp^{-1}, corresponding to a microscopic binding constant of about 14 M^{-1} bp^{-1}. A variety of single mismatches were introduced in 9- and 12-mer PNA−DNA duplexes. Melting temperatures (T_m) of 9- and 12-mer PNA−DNA duplexes with a single mismatch dropped typically 15−20 °C relative to that of the perfectly matched sequence with a corresponding free energy penalty of about 15 kJ mol^{-1} bp^{-1}. The average cost of a single mismatch is therefore estimated to be on the order of or larger than the gain of two matched base pairs, resulting in an apparent binding constant of only 0.02 M^{-1} per mismatch. The impact of a mismatch was found to be dependent on the neighboring base pairs. To a first approximation, increasing the stability of the surrounding region, i.e., the distribution of A·T and G·C base pairs, decreases the effect of the introduced mismatch.

Peptide nucleic acids (PNAs)[1] are DNA mimics with interesting potential for gene therapeutic, gene diagnostic, and molecular biology applications (*1*, *2*). Compared to other DNA analogues, PNA has several advantages, mainly arising from the fact that the negatively charged backbone of DNA is replaced by a neutral pseudopeptide in PNA. The backbone in PNA is composed of *N*-(2-aminoethyl)glycine units, onto which the natural recognizing elements of DNA (the nucleobases) are attached (Figure 1).

Designed to be structurally homomorphous to DNA, a mixed-sequence PNA strand is capable of hybridizing with complementary single-stranded DNA, RNA, or another PNA strand to form highly stable duplexes with a high degree of sequence selectivity (*3*−*6*). At moderate salt levels, the thermal stabilities increase, mainly due to the lack of interstrand electrostatic repulsion in PNA-containing duplexes (*4*, *6*), in the following order: DNA−DNA < PNA−DNA < PNA−RNA < PNA−PNA. The structures adopted by duplexes of PNA−nucleic acid hybrids can be characterized as having Watson−Crick type base pairing and stacking patterns similar, though not identical, to those of DNA−DNA duplexes (*7*). The PNA−DNA duplex is somewhat unwound (helical rise of 42 Å with 13 bp per turn, and a slightly larger helix diameter of 23 Å) with a major groove wider and a minor groove narrower than those of the DNA−DNA duplex (*7*).

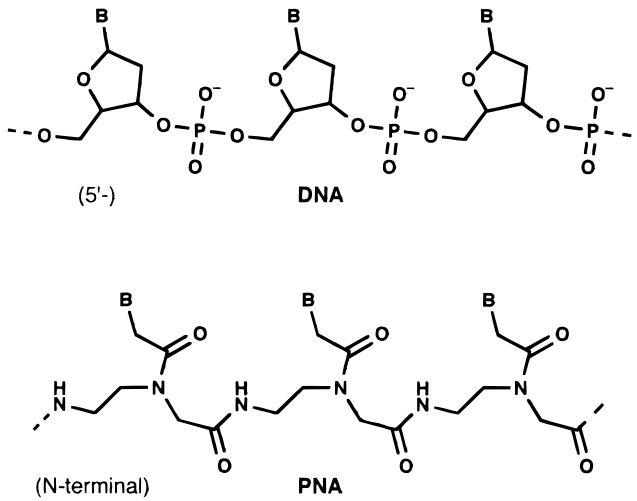

FIGURE 1: Chemical structures of DNA and PNA. The deoxyribose phosphodiester backbone in DNA has been changed to *N*-(2-aminoethyl)glycine in PNA.

In addition, PNAs have resistance to nuclease and proteolytic degradation, and may thus be developed into gene therapeutic antisense and antigene drugs (*1*, *2*, *8*). Furthermore, PNA may replace DNA oligonucleotides in several biotechnology and molecular biology hybridization applica-

† Financial support from the Swedish Cancer Foundation (B.N.), the EU Biomed 2 Program (BMH4-CT96-0848 and BMH4-CT96-1081, to B.N. and P.E.N.), and the Danish National Research Foundation (P.E.N.) is gratefully acknowledged.

* To whom correspondence should be addressed. E-mail: tmr@phc.chalmers.se. Phone: +46-31-772 30 49. Fax: +46-31-772 38 58.

‡ Chalmers University of Technology.

§ AstraZeneca R&D Mölndal.

|| The Panum Institute.

[1] Abbreviations: PNA, peptide nucleic acid; FET, fluorescence energy transfer; ABS, absorption (hypochromicity); ITC, isothermal titration calorimetry; DSC, differential scanning calorimetry; CD, circular dichroism.

10.1021/bi000039g CCC: $19.00
Published on Web 06/06/2000

tions such as in situ hybridization (*9*), hybridization arrays (*10*), and hybridization sensors (*11*). Finally, many novel applications have been developed on the basis of the unique properties of PNA. These include, for instance, PCR clamping (*12*), rare cleavage strategies for genome analyses (*13*), and sample preparation (*14*).

We recently studied the thermodynamics of PNA hybridization, combining different spectroscopic methods (FET and ABS) with ITC, and found that the PNA−DNA hybridization could be reasonably well described by a simple two-state (all-or-none) binding model (*15*). Although there were no significant indications of precursor complexes, the binding characterization was complicated by relatively strong interactions between the incorporated labels and the duplexes for fluorescence detection (*15*).

Our previous study was limited to a single decameric sequence (*15*). Here we continue with the determination of the strength of binding of PNA to DNA and, first, study the length dependence (6−20-mers) of the binding thermodynamics using mainly absorption spectroscopy. The average binding constant was determined to be about 14 M^{-1} bp^{-1}. The sequence dependence of PNA−DNA duplexes is most likely different from that of the DNA−DNA duplex due to the asymmetry of PNA−DNA duplexes (*16*). Whereas nearest-neighbor data for DNA−DNA duplexes have been shown to describe the hybridization thermodynamics satisfactorily (*17*), this inherent asymmetry requires additional parameters if sequence-dependent melting temperatures are to be predicted, as pointed out by Giesen et al. (*18*). The asymmetry of PNA−DNA duplexes would further necessitate the use of a full set of nearest-neighbor parameters ($2^4 = 16$). We now report a thermodynamic study of the sequence specificity of the PNA binding to DNA. A number of mismatched 9- and 12-mer PNA−DNA duplex series were subject to thermal denaturation analysis, yielding an average single mismatch binding constant of about 0.02 M^{-1}, i.e., lower by a factor of about 600 than that of the fully matched base pair duplex. Our data on the different mismatches will be compared with published PNA−DNA data (*5*, *19*) as well as with DNA−DNA nearest-neighbor data (*17*) on mismatched duplexes.

MATERIALS AND METHODS

Chemicals. Mixed-sequence strands of PNA (*20*) and DNA were studied; base sequences and notations are given in Figure 2. Melting curves were recorded in phosphate buffer (pH 7, 100 mM NaCl) with EDTA, except for the mismatch experiments which were carried out in 10 mM sodium phosphate buffer (pH 7.7).

Concentration Determinations. Strand concentration determinations were based on absorption (at 260 nm) measured at 80 °C on a Cary 4 spectrophotometer equipped with a programmable multicell temperature block. At 80 °C, the nucleobases are considered completely destacked and the absorptivity is therefore assumed to be the sum of the absorptivities of DNA monomers taken from the literature (*21*).

Hybridizations and Melting Experiments. The appropriate single strands were hybridized by the following procedure: heating of the solutions of single strands to 80 °C, adjustment of concentration by absorbance measurements, and mixing

(a) Sequences included in fit $\Delta G°$ vs N

1	H-ac-TTTTGG-Lys-NH_2
2	H-ATCAAC-Lys-NH_2
3	H-GTAGATCA-NH_2
4	H-GCACAGCC-Lys-NH_2
5	H-TCACTACTA-Lys-NH_2
6	H-TCACAACTA-Lys-NH_2
7	H-TCACCACTA-Lys-NH_2
8	H-TCACGACTA-Lys-NH_2
9	H-GTACATCACT-Lys-NH_2
10	H-ACTGATCTAC-NH_2
11	H-GCATATCACC-NH_2
12	H-CATCTAGTGA-Lys-NH_2
13	H-GTGTGGTTTT-Lys-NH_2
14	H-TTTTGGTGTG-Lys-NH_2
15	H-GTAGATCACT-Lys-NH_2
16	H-AGTGATCTAC-Lys-NH_2
17	H-GTAGATCACTGT-NH_2
18	H-CCCCTATTGTCC-Lys-NH_2
19	H-TTCTGCCTTTCT-Lys-NH_2
20	H-aha-TATCATTGGTGT-NH_2
21	H-ACGTCACAACTA-Lys-NH_2
22	H-ACAGTGATCTAC-Lys-NH_2
23	H-aha-ACACCAATGATA-NH_2
24	H-Gly-TTACCCTCGATCG-Lys-NH_2
25	H-TGTACGTCACAACTA-NH_2
26	H-aha-TCAAGACGCATGTCT-NH_2
27	H-ac-ACGCCACATCTTGCG-D-Lys-NH_2
28	H-CTTTTTGTCGACGACC-Lys-NH_2
29	H-Gly-TCTCCATCCTCTTCACT-NH_2
30	H-TATTCCGTCATCGCTCCTCA-Lys-NH_2
31	H-ac-GTAACACGCCACATCTTGCG-D-Lys-NH_2

Sequences not included in fit $\Delta G°$ vs N

32	H-ac-CCACACCGA-NH_2
33	H-TTATCATCTC-Lys-NH_2
34	H-GTAGTTCACT-Lys-NH_2
35	H-Gly-AAAAAAAAAA-NH_2
36	H-AAATGGTGAG-Lys-NH_2
37	H-AAAAGGAGAG-Lys-NH_2
38	H-TTTTCCTATATT-Lys-NH_2
39	H-aha-ATCATCTTTGGT-NH_2
40	H-Gly-TCCCTCGTAAACTCT-NH_2
41	H-TAGTTATCTCTATCT-Lys-NH_2
42	H-aha-CTGCAGCTTCCTCAT-NH_2
43	H-aha-TCGCGTTGTTCGTGA-NH_2
44	H-GAGTGTTAAGGTGTG-Lys-NH_2

(b) Mismatched sequences included in Figure 6

6 (X=T)			
45 (X=A)	H-TCACAACTA-Lys-NH_2		P9
46 (X=C)	3'-AGTG**X**TGAT-5'		D9A**X**
47 (X=G)			
6 (X=A)			
48 (X=C)	H-TCAC**X**ACTA-Lys-NH_2		P9**X**T
49 (X=G)	3'-AGTGTTGAT-5'		D9
50 (X=T)			
21 (X=G)			
51 (X=A)	H-ACGTCACAACTA-Lys-NH_2		P12
52 (X=C)	3'-TGCAGT**X**TTGAT-5'		D12C**X**
53 (X=T)			

Mismatched sequences not included in Figure 6

6 (X=G)			
54 (X=A)	H-TCACAACTA-Lys-NH_2		P9
55 (X=C)	3'-AGT**X**TTGAT-5'		D9C**X**
56 (X=T)			

FIGURE 2: (a) Numbering and sequences of PNA strands of the studied PNA−DNA duplexes referenced in fitting linear approximations to data shown in Figures 3 and 6. Abbreviations: Lys, L-lysine (D-configuration denoted specifically); aha, 6-aminohexanoic acid; Gly, L-glycine; ac, acetyl. (b) Numbering, sequences, and notations of the series of mismatched PNA−DNA duplexes that were studied. Strands positioned in antiparallel orientation. The mismatches are underlined and denoted starting with the base on the PNA strand throughout.

(at 80 °C) of the single strand solutions, followed by slow cooling to the starting temperature (2−4 °C) of the melting experiments. To avoid moisture on the cuvettes, the sample compartment of the spectrophotometer was purged with dry air.

CD Measurements. Circular dichroism is defined as the difference between the absorption of left, A_l, and right, A_r, circularly polarized light (*22*).

$$\mathrm{CD} = A_l - A_r \quad (1)$$

CD measurements were taken on a Jasco 700 spectropolarimeter.

Evaluation of Melting Curves. The absorbance at 260 nm, plotted versus temperature, was fitted with a six-parameter function to a two-state model (*23*) of the hybridization. This model assumes that single strands are in equilibrium with only one base-paired duplex structure; i.e., there are no partially base-paired structures in the melting process (*24*). Furthermore, the thermodynamic parameters $\Delta H°$ and $\Delta S°$ are assumed to be temperature-independent, and the change in heat capacity (ΔC_p) between the two states is assumed to be zero. These assumptions might be an oversimplification as reported recently by Rouzina et al. (*25*, *26*), but since we here deal mostly with (relative) changes in $\Delta G°$ and T_m, we use this simplest of models in our analysis.

The equilibrium constant K can be expressed in either of the following exponential forms of eq 2 (depending on which pair of thermodynamic parameters are fitted, $\Delta H°$ and $\Delta S°$ or $\Delta H°$ and T_m) as well as in terms of the extent of reaction α (fraction of strands in duplex form).

$$K = \exp\left(-\frac{\Delta H°}{RT} + \frac{\Delta S°}{R}\right) = \frac{4}{c_T/c°}\exp\left(-\frac{\Delta H°}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right)\right) = \frac{2\alpha}{(1-\alpha)^2 c_T/c°} \quad (2)$$

where c_T is the total concentration of single strands and $c°$ is the unit concentration (1 M). Solving for α gives the expression $\alpha = f(\Delta H°, \Delta S°, T)$ or $\alpha = g(\Delta H°, T_m, T)$ which is combined with a model in which the physical property (e.g., A_{260}) of pure single- and double-stranded forms are allowed to vary with temperature according to a linear approximation, $A_s = m_s T + b_s$ and $A_d = m_d T + b_d$, yielding a fitting equation of the form

$$A = \alpha A_d + (1-\alpha) A_s \quad (3)$$

containing a total of six fitting parameters: $\Delta H°$, T_m, m_d, b_d, m_s, and b_s (*24*, *27*, *28*). The deviations of m_s and m_d from zero can be regarded as a consideration of the temperature dependence of nucleobase stacking dynamics.

Isothermal Titration Calorimetry. A MicroCal ITC MC-2 system was used in conjunction with an electronically controlled circulating water bath. A solution of one of the strands (about 5 μM) was placed in the cell (volume of 1.4 mL) and the titrant solution (about 0.1 mM) in a 100 μL syringe, whose needle was designed as a paddle-shaped stirrer rotating at 400 rpm. The syringe is controlled by a stepping motor, allowing precise injections (±0.1 μL). Typically, 20 injections of 4 μL each and 5 min apart were made. The integrated peaks (pulses) of the heat production upon each injection were plotted against the molar ratio. With the built-in software, MicroCal Origin, the binding isotherms were fitted to a two-state binding model, giving numerical values of both the enthalpy of binding (ΔH) and the binding constant (K).

RESULTS

Length Dependence of the Binding Free Energy for Perfectly Matched PNA−DNA Duplexes. To investigate the average base pair contribution (as $\Delta G°$/bp) to the strength of binding of PNA to DNA strands, melting curves monitored by absorbance at 260 nm were analyzed by fitting to eqs 2 and 3 with $\Delta H°$ and T_m as fitting parameters, as described in Materials and Methods. A total of 44 fully complementary duplexes of different lengths N (ranging from 6- to 20-mers) were investigated (Figure 2a). Analysis of the melting curves gave binding free energies $\Delta G°$ that followed an approximately linear relationship with respect to the number of base pairs. A linear fit of $\Delta G°$ versus N was made, including the first 31 listed mixed sequences in Figure 2a, while excluding sequences 32−44 mainly due to their G·C content (<35 or >65%) and/or very low (<20%) or very high (>80%) fractional pyrimidine content (on PNA). Figure 3a shows data ($\Delta G°$ vs N, and K vs N) with the average contribution per base pair ($\partial\Delta G°/\partial N$) determined to be -6.5 ± 0.3 kJ mol^{-1} bp^{-1}. Converting the observed $\partial\Delta G°/\partial N$ value into the corresponding average binding constant for formation of one base pair (k) yields (at 25 °C) a value of $\sim 14 \pm 1$ M^{-1}. In Figure 3b, the free energy difference is resolved into enthalpic and entropic terms, with average slopes of -30.0 ± 2.5 kJ mol^{-1} bp^{-1} ($\partial\Delta H°/\partial N$) and -23.5 ± 2.3 kJ mol^{-1} bp^{-1} ($\partial T\Delta S°/\partial N$), respectively, both with intercepts close to zero at zero length.

UV Melting Curves on 9-mer Single-Mismatch Series. The recorded absorbance data for the 9-mer series P9/D9A$\underline{\mathrm{X}}$ (where X = T, A, C, or G; see Figure 2b for abbreviations and notations) shown in Figure 4a were fitted to eqs 2 and 3, as described in Materials and Methods. In these duplexes, the central A·$\underline{\mathrm{T}}$ base pair is mismatched (variations made on the DNA strand). Figure 4 shows that the three mismatched duplexes have melting temperatures ($\Delta T_m \approx -15$ °C) considerably lower than that of the fully complementary duplex (P9/D9A$\underline{\mathrm{T}}$). The mismatches with the largest penalty in duplex stability are the two purine·purine base pairs (A·$\underline{\mathrm{A}}$ and A·$\underline{\mathrm{G}}$). Fitted thermodynamic parameters for association of single strands with duplexes are shown graphically in Figure 4b. The binding free energy of the mismatches is around 11 kJ mol^{-1} less favorable, i.e., $\Delta\Delta G°(25\ °C) \approx 11$ kJ mol^{-1}, than that for the perfectly sequence-matched duplex (Table 1).

When changing the mismatch to the PNA strand with the same central $\underline{\mathrm{A}}$·T base pair mismatched (P9$\underline{\mathrm{X}}$T/D9; see Figure 2b), we recorded somewhat less defined melting curves (with T_m values just above 10 °C) for the cases with $\underline{\mathrm{C}}$·T and $\underline{\mathrm{T}}$·T mismatches, and with on average $\Delta\Delta G°$(25 °C) being ≈17 kJ mol^{-1} relative to the fully matched duplex (Table 1). The T_m of the $\underline{\mathrm{G}}$·T mismatch, however, is only 8 °C lower and with $\Delta\Delta G°$(25 °C) being ≈5 kJ mol^{-1} in comparison with the perfectly matched duplex (Table 1).

For the same 9-mer sequence, we performed further melting curve analysis, but with a C·$\underline{\mathrm{G}}$ base pair mismatched

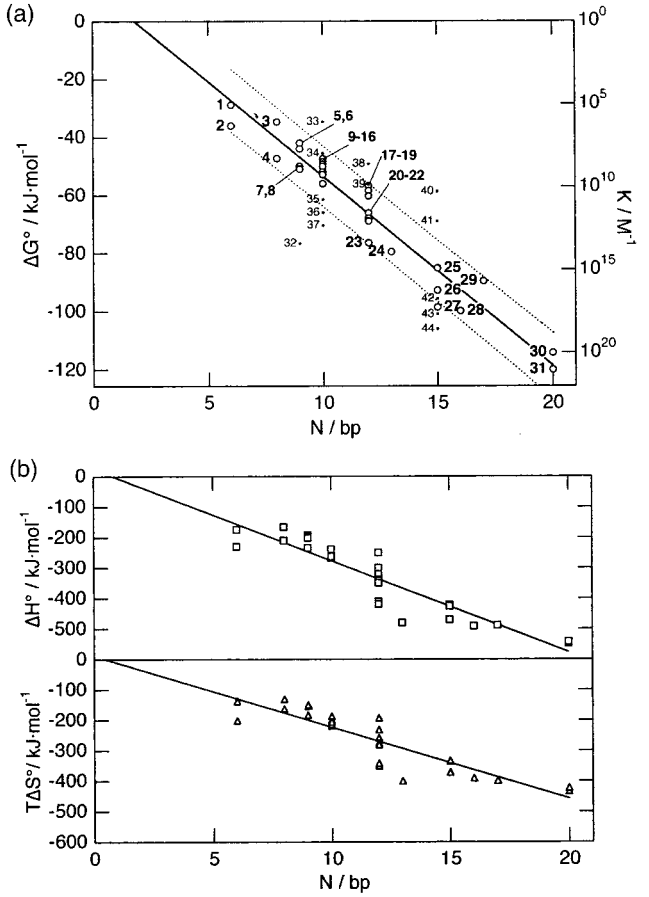

FIGURE 3: (a) Length dependence (base pairs, N) of PNA−DNA duplex stability for matched sequences, given in terms of binding free energy ($\Delta G°$, left) and binding constant (K, right logarithmic scale). All data points are numbered according to Figure 2a, and sorted from top left to bottom right (increasing N and decreasing $\Delta G°$). Only points 1−31 (○) were included in the linear fit, whereas points 32−44 (•) were omitted as outliers from our selected data set of "mixed" sequences. The linear fit yields $\Delta G° = (-6.5 \pm 0.3)N + (12 \pm 4)$ kJ mol^{-1} and the corresponding $K = (14 \pm 1)^N$ M^{-1}. The dotted curves show the 90% prediction interval (model plus random noise) for the $\Delta G°$ vs N fit. (b) Length dependence of enthalpy $\Delta H°$ (top) and entropy $T\Delta S°$ (bottom) of formation of PNA−DNA duplexes for matched sequences (only points 1−31 from Figure 2a shown, as selected in the $\Delta G°$ vs N plot). The linear fits yield $\Delta H° = -(30.0 \pm 2.5)N + (23 \pm 30)$ kJ mol^{-1} and $T\Delta S° = -(23.5 \pm 2.3)N + (11 \pm 28)$ J mol^{-1} (at 25 °C).

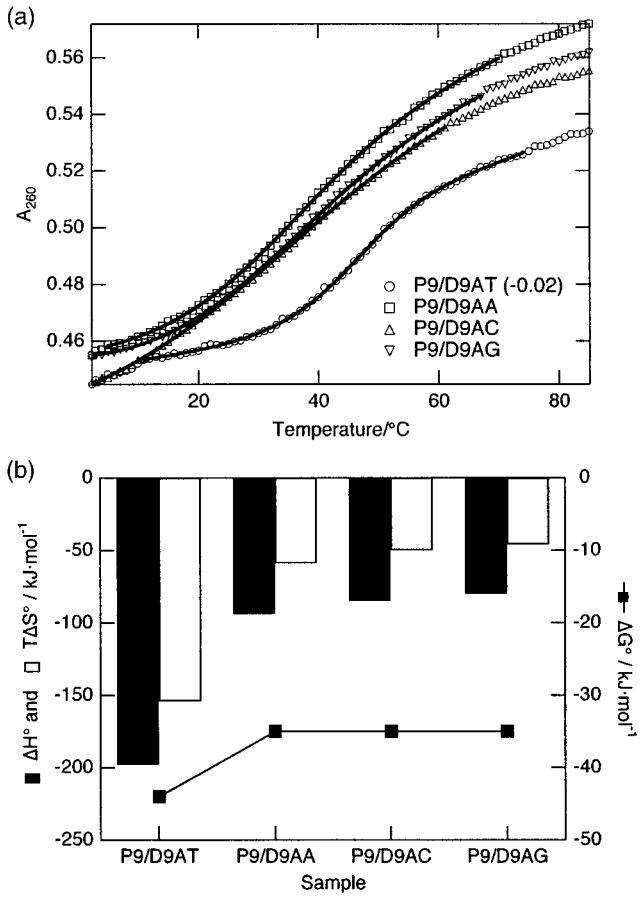

FIGURE 4: (a) Absorbance melting curves of the 9-mer PNA−DNA system (P9/D9AX) with a central single mismatch (from the A•T base pair). The curve for the perfectly matched sequence (P9/D9AT) is shifted (−0.02 unit) for clarity. Lines show the fitted function according to eqs 2 and 3. (b) Corresponding fitted thermodynamic parameters ($\Delta H°$, $T\Delta S°$ to the left, and $\Delta G°$ to the right) for the formation of duplexes from single strands of the same systems as described above.

by variation in the DNA strand (P9/D9CX; see Figure 2b). This resulted in poorly defined melting curves with T_m values around 0 °C and too low to measure accurately, but estimations of their respective ΔT_m values are included in Table 2 for comparative purposes.

UV Melting Curves on 12-mer Mismatch Series. Melting curves of the 12-mer series P12/D12CX (Figure 2b) shown in Figure 5a were fitted to eqs 2 and 3 as previously described in Materials and Methods. The sequence is related to the 9-mer by having a three-base extension at the 3′-terminus of the DNA strand. In these duplexes, a central C•G base pair is mismatched (on the DNA strand). As seen in Table 1, the mismatched 12-mer duplexes have melting temperatures ($\Delta T_m \approx -19$ °C) significantly lower than that of the fully complementary duplex (P12/D12CG). Values of the fitted thermodynamic parameters for association of single strands to form duplexes are shown in Figure 5b. On average, the binding free energy of the mismatched duplexes is some 25 kJ mol^{-1} less favorable [$\Delta\Delta G°(25$ °C) ≈ 25 kJ mol^{-1}] than that for the perfectly complementary duplex (Table 1).

Table 1: Thermodynamic Parameters of the Formation of the Studied PNA−DNA Duplexes from Single Strands, Obtained from Absorbance Melting Curves[a]

sample	T_m[b] (°C)	$\Delta H°$ (T_m) (kJ mol^{-1})	$T\Delta S°$ (298 K) (kJ mol^{-1})	$\Delta G°$ (298 K) (kJ mol^{-1})
9-mers (H-TCACAACTA/3′-AGTGXTGAT)[c]				
P9/D9AT	42	−198	−154	−44
P9/D9AA	28	−94	−59	−35
P9/D9AC	29	−85	−50	−35
P9/D9AG	25	−80	−46	−35
9-mers (H-TCACXACTA/3′-AGTGTTGAT)[c]				
P9AT/D9	42	−198	−154	−44
P9CT/D9	10	−123	−96	−27
P9GT/D9	34	−170	−131	−39
P9TT/D9	13	−127	−99	−28
12-mers (H-ACGTCACAACTA/3′-TGCAGTXTTGAT)[d]				
P12/D12CG	62 (64)[d]	−300	−232	−68
P12/D12CA	43 (45)[d]	−144	−102	−43
P12/D12CC	41 (43)[d]	−115	−74	−40
P12/D12CT	47 (48)[d]	−145	−101	−44

[a] Measured in 10 mM phosphate buffer (pH 7.7). [b] Uncertainties in T_m values of less than ±1 °C. [c] At a strand concentration of 5 μM. [d] At a strand concentration of 3.8 μM. Values of T_m in parentheses correspond to 5 μM.

Table 2: Thermal and Thermodynamic Cost of Various Mismatches of the Studied PNA−DNA Duplexes in Comparison with Earlier Studies

Mismatch	Sequence	ΔT_m (°C)	K/K_{mm}	*Ref.*
A•A	H-TCACAACTA 3′-AGTGATGAT	−14	47	
	H-TGTACGTCACAACTA 3′-ACATGCAGAGTTGAT	−15.0	100	a
	H-TATCATTGGTG 3′-TATAGAAACCACAAAGG	−17.6	29	b
A•C	H-TCACAACTA 3′-AGTGCTGAT	−13	51	
	H-TGTACGTCACAACTA 3′-ACATGCAGCGTTGAT	−13.5	16	a
	H-TATCATTGGTG 3′-TATAGCAACCACAAAGG	−19.4	n.a.	b
A•G	H-TCACAACTA 3′-AGTGGTGAT	−17	74	
	H-TGTACGTCACAACTA 3′-ACATGCAGGGTTGAT	−12.5	13	a
	H-TATCATTGGTG 3′-TATAGGAACCACAAAGG	−16.6	173	b
C•T, C•T	H-TCACCACTA 3′-AGTGTTGAT	−32	1×10^3	
	H-TCACAACTA 3′-AGTTTTGAT	< −45	n.a.	
	H-ACGTCACAACTA 3′-TGCAGTTTTGAT	−15	13×10^3	
	H-TGTACGTCACAACTA 3′-ACATGCATTGTTGAT	−17.5	0.26×10^3	a
	H-TATCATTGGTG 3′-TATATTAACCACAAAGG	−17.7	n.a.	b
G•T, G•T	H-TCACGACTA 3′-AGTGTTGAT	−8	10	
	H-TGTACGTCACAACTA 3′-ACATGTAGTGTTGAT	−8.0	22	a
	H-TATCATTGGTG 3′-TATAGTAATCACAAAGG	−13.9	n.a.	b
	H-TATCATTGGTG 3′-TATAGTAACTACAAAGG	−12.7	n.a.	b
T•T, T•T	H-TCACTACTA 3′-AGTGTTGAT	−29	0.64×10^3	
	H-TGTACGTCACAACTA 3′-ACATGCTGTGTTGAT	−10.0	4.3	a
	H-TATCATTGGTG 3′-TATTGTAACCACAAAGG	−5.2	19	b
	H-TATCATTGGTG 3′-TATAGTTACCACAAAGG	−9.3	38	b
	H-TATCATTGGTG 3′-TATAGTATCCACAAAGG	−7.9	22	b
C•A	H-TCACAACTA 3′-AGTATTGAT	< −45	n.a.	
	H-ACGTCACAACTA 3′-TGCAGTATTGAT	−19	24×10^3	
	H-TGTACGTCACAACTA 3′-ACATGCAATGTTGAT	−18.5	0.16×10^3	a
	H-TATCATTGGTG 3′-TATAATAACCACAAAGG	−18.9	n.a.	b
C•C	H-TCACAACTA 3′-AGTCTTGAT	< −45	n.a.	
	H-ACGTCACAACTA 3′-TGCAGTCTTGAT	−21	64×10^3	
	H-TGTACGTCACAACTA 3′-ACATGCACTGTTGAT	−19.5	0.63×10^3	a
	H-TATCATTGGTG 3′-TATACTAACCACAAAGG	−21.0	1.4×10^3	b

[a] Data from Kilså Jensen et al. (*5*). [b] Data from Igloi (*19*).

DISCUSSION

Length Dependence of the Binding Free Energy for Perfectly Matched PNA−DNA Duplexes. One important issue to investigate for potential application of PNA to single-stranded DNA targeting, relevant especially for diagnostic applications, is the average contribution per base pair to the binding of single-stranded PNA to single-stranded DNA and whether it can be generalized to predict stability for various sequences. Our melting curve analysis for different lengths N (from 6- to 20-mer) of fully complementary mixed-sequence PNA−DNA duplexes yielded an average contribution of -6.5 ± 0.3 kJ mol^{-1} bp^{-1} ($\partial\Delta G°/\partial N$), corresponding to a binding constant for one base pair (k) of about 14 M^{-1} (at 25 °C, Figure 3a).

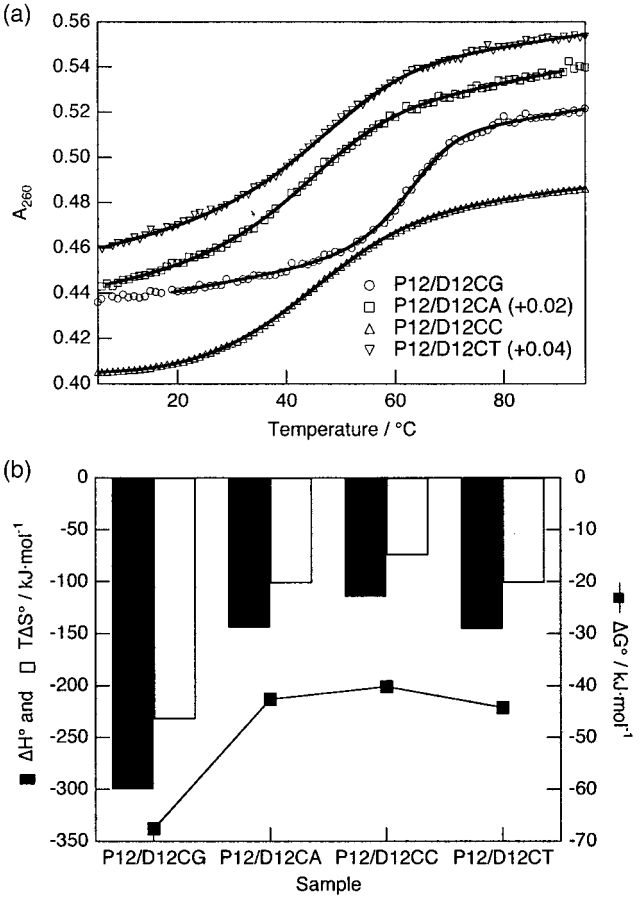

FIGURE 5: (a) Absorbance melting curves of the 12-mer PNA−DNA system (P12/D12CX) with a single mismatch (from the C•G base pair). Two of the curves are shifted for clarity. Lines show the fitted function according to eqs 2 and 3. (b) Corresponding fitted thermodynamic parameters ($\Delta H°$, $T\Delta S°$ to the left, and $\Delta G°$ to the right) for the formation of duplexes from single strands of the same 12-mer systems as described above.

Compared to nearest-neighbor parameters (*29*) for the formation of DNA−DNA base pairs, the length-averaged enthalpy and entropy changes we obtain for the present data set of mixed sequences (Figure 3b) are of reasonable magnitude; e.g., the average $\Delta H°$ for the 10 nearest neighbors in the DNA−DNA duplex is −33 kJ mol^{-1} bp^{-1}, and $T\Delta S°$ is −26 kJ mol^{-1} bp^{-1} at 25 °C, thus yielding a $\Delta G°$ of −7 kJ mol^{-1} bp^{-1}. This only speciously indicates that the DNA−DNA duplex on average is more stable than the PNA−DNA duplex, since the nearest-neighbor data are given at 1 M NaCl whereas our data are valid under low-salt conditions (10−100 mM), making the comparison unfavorable by some 20−30 °C in T_m for a model mixed-sequence 10-mer (*6*). This decrease in T_m corresponds roughly to a less favorable binding free energy of 3−4 kJ mol^{-1} bp^{-1} at 25 °C (*17*) and would thus give a better estimate of the parameter $\partial\Delta G°/\partial N$ for the DNA−DNA duplex of −3 to −4 kJ mol^{-1} bp^{-1}, in good accordance with $^1/_{10}$ of the value we found previously for the 10-mer model DNA−DNA duplex in low-salt (10 mM) buffer (*15*). Converting this to a binding constant k for one average base pair gives a value of approximately 4 M^{-1}, i.e., less than $^1/_3$ of that of PNA−DNA duplex formation. The difference between PNA−DNA and DNA−DNA binding strength at physiological conditions will be slightly smaller, but it still makes it possible to use much shorter PNA sequences to achieve high-affinity binding.

An interesting observation that calls for comment relates to points 20 and 23 in Figure 3a, corresponding to two 12-mer duplexes where the PNA strands are H-aha-TATCAT-TGGTGT-NH_2 and H-aha-ACACCAATGATA-NH_2, respectively. These two duplexes have the same sequence of base pairs (G·C content of 33%) but with the backbones interchanged. It turns out that even though these duplexes are virtually identical in terms of sequence, their binding free energies $\Delta G°$ are significantly different. Whereas the former duplex has a $\Delta G°$ of -68 kJ mol^{-1} corresponding to a binding constant K of $\approx 8 \times 10^{11}$ M^{-1}, the latter has a $\Delta G°$ of -78 kJ mol^{-1} and a K of $\approx 5 \times 10^{13}$ M^{-1}. Averaged over the number of base pairs, these values give rise to average binding constants for one base pair of about 10 and 14 M^{-1}, respectively, so it appears that placing the G's on the DNA strand gives the most favorable $\Delta G°$ of the duplex. Inspecting the sequences in more detail, and using the set of nearest-neighbor data for DNA−DNA and PNA−DNA-specific corrections from Giesen et al. (*18*), reveals that the former 12-mer PNA−DNA duplex is expected to have a substantially lower thermal stability with T_m lowering contributions from both DNA−DNA nearest-neighbor data and the fractional pyrimidine content of the PNA strand. However, we observe an only 2 °C difference in T_m, while their empirically determined formula (*18*) predicts a ΔT_m as large as 11 °C. What is interesting to note and to remember when analyzing melting curves is that even though the thermal stabilities are similar, the thermodynamic stability might differ significantly as in the present example. The binding free energy approach has been used to predict the thermodynamic properties of PNA−DNA duplexes (*30*), although the authors did not include any correction for the inherent asymmetry in PNA−DNA duplexes. Using the data from Giesen et al. (*18*) in terms of the binding constant per base pair (k), the DNA−DNA nearest-neighbor data account for approximately half the difference between our values of 10 and 14 M^{-1}, respectively. The other part is then attributed to an effect of the pyrimidine content of the PNA strand.

Effects of Mismatches on the Thermal Stability (T_m) of PNA−DNA Duplexes. We have investigated four different series of mismatched PNA−DNA duplexes (see Figure 2b). Two of them differ only in the chemical nature of the two strands, i.e., the 9-mer systems P9/D9A$\underline{X}$ and P9$\underline{X}$T/D9. Whereas the P9/D9A$\underline{X}$ series has a central A base on the PNA strand and the opposing X base mismatched, the series P9$\underline{X}$T/D9 contains the mismatch (X) on the PNA strand opposite a T base on the DNA. Thus, both these mismatched series originate from the same sequence with a perfectly matched central A·T base pair, and provide us with the mismatches (the base on the PNA strand given first throughout, and the mismatched base is underlined to show from which base pair, i.e., A·T or G·C, the mismatch series originates) A·$\underline{A}$, A·$\underline{C}$, A·$\underline{G}$, $\underline{C}$·T, $\underline{G}$·T, and $\underline{T}$·T. As our results in Table 1 show, disruption of this A·T base pair has greatly varying impact on duplex stability. In addition, we have investigated two systems (a 9-mer system, P9/D9C$\underline{X}$, and a 12-mer system, P12/D12C$\underline{X}$; see Figure 2b) with a mismatch originating from a G·C base pair. These supply us with mismatches C·$\underline{A}$, C·$\underline{C}$, and C·$\underline{T}$ and will be discussed briefly at the end of this section. Summarized in Table 2 are also data taken from two previous studies where PNA−DNA mismatches were investigated. (i) The first is a study of hybridization kinetics by Kilså Jensen et al. (*5*) of a 15-mer duplex with some of the same sequence as our 9-mer (H-TGTACGTCACAACTA-NH_2) (ii) Using affinity electrophoresis, Igloi (*19*) studied the hybridization (onto immobilized PNAs) of 11-mer PNA−DNA duplexes.

In the P9/D9A$\underline{X}$ series, melting temperatures decrease considerably (on average, $\Delta T_m \approx -15$ °C), as can be seen in Figure 4 and Table 1. The two purine·purine base pairs (A·$\underline{A}$ and A·$\underline{G}$) cost more than the A·$\underline{C}$ mismatch which induces the least destabilization (T_m values A·$\underline{C}$ > A·$\underline{A}$ > A·$\underline{G}$). This is somewhat surprising since the stabilization conferred by stacking of the inherently larger purine bases might be expected to be greater than that of the pyrimidine cytosine, as shown for DNA−DNA (*17*) as well as for PNA−PNA duplexes (*7*) with T_m values declining in the following order: A·$\underline{A}$ > A·$\underline{G}$ > A·$\underline{C}$. Kilså Jensen et al. (*5*) found a different relative order (A·$\underline{G}$ > A·$\underline{C}$ > A·$\underline{A}$) for PNA−DNA mismatches, but for PNA−RNA duplexes, the A·$\underline{C}$ mismatch was the most stable. The stability order found by Igloi (*19*) was as follows: A·$\underline{G}$ > A·$\underline{A}$ > A·$\underline{C}$. Even though we find a relative order of T_m values for A·$\underline{X}$ mismatches different from that reported in two previous PNA−DNA studies (*5*, *19*), the differences between T_m values are small within the studies (<4 °C), and change of the order could possibly lie within experimental error in T_m determinations.

For the P9$\underline{X}$T/D9 series, the $\underline{G}$·T mismatch was the least destabilizing, with only a modest (8 °C) decrease in T_m (Table 1), the same as Kilså Jensen et al. (*5*) found for the two most stable mismatches, $\underline{G}$·T (Table 2) and T·$\underline{G}$. Igloi (*19*) found the $\underline{G}$·T mismatch to be one of the most stable (Table 2, $\Delta T_m = -13.3$ °C on average). In DNA−DNA systems, the $\underline{G}$·T mismatch is also found to be extraordinarily stable and is believed to achieve this by forming a wobble base pair with two hydrogen bonds (*31*). In stark contrast, we observe dramatic drops of around 30 °C in T_m for both the $\underline{C}$·T and $\underline{T}$·T mismatches, as will be discussed in the following section.

To check the surprisingly low T_m values found for these mismatches, we measured melting curves for PNA−DNA duplexes where the respective mismatch (on PNA strand) was compensated by using a DNA fully complementary to the previously mismatched PNA. Melting temperatures as well as thermodynamic parameters for all these duplexes were close to those of the original P9$\underline{A}$T/D9 duplex, indicating that there is nothing intrinsically unstable about these sequences. The P9$\underline{T}$A/D9′ duplex (where the prime denotes that this DNA has a sequence differing one base from that of the original DNA) was found to have a T_m of $\approx$38 °C and a $\Delta G°$ of ≈ -42 kJ/mol which is only slightly lower than that of the original P9$\underline{A}$T/D9. In contrast, the matched duplexes P9$\underline{C}$G/D9′ and P9$\underline{G}$C/D9′ were found to have thermal (T_m values of 47 and 53 °C) and thermodynamic stabilities ($\Delta G°$ values of -50 and -51 kJ mol^{-1}) significantly higher than those of the parent duplex, in good agreement with what is expected for the 9-mers containing one additional (now four) G·C base pair in their sequences. These data points are also included in our data set for fully matched duplexes, e.g., for the length dependence of the binding free energy (Figure 3).

Since the T_m of the P9$\underline{C}$T/D9 duplex is quite low and close to the lower limit of experimental temperatures, the observed

decrease in T_m (32 °C) is to be considered less reliable. To obtain another estimate of the thermal cost of this mismatch, the 12-mer duplex P12/D12CT was also studied and gave a ΔT_m of −15 °C, a drop only slightly smaller than the results found by Kilså Jensen et al. (*5*) and Igloi (*19*) (see Table 2).

The most aberrant result from our observations, compared to the two other PNA−DNA mismatch studies (*5, 19*), is that the T·T mismatch is unstable relative to G·T. Igloi (*19*) concluded that T·T mismatches are the most stable (ΔT_m = −7.5 °C on average), and in the work by Kilså Jensen et al. (*5*), T·T was found to be one of the most stable mismatches (ΔT_m = −10 °C). We checked the stability of the series P9XT/D9 duplexes with ITC (data not shown), and obtained results consistent with those from melting curves. The same interpretation was made from CD spectra for these mismatches (see the Supporting Information). Due to simplistic stacking (size) considerations, one could argue that the pyrimidine T·T mismatch is less stable than the G·T mismatch, and in general, we find that the most stable mismatches are when purines are replaced with purines or pyrimidines with pyrimidines. In a PNA−PNA mismatch assay, it was found that the T·T mismatch was much more costly than both G·T and T·G in terms of duplex stability (*7*). Even though we cannot specifically explain this behavior of the T·T mismatch, these examples lead us to think that the loss of stacking for the small T·T step in the structurally wider PNA-containing duplexes might be greater than in the corresponding DNA−DNA duplexes, possibly originating from the different nature and flexibility of the backbones.

Context dependence has been shown for DNA−DNA duplex mismatches previously (*17*) and was recently completed to a full set of nearest-neighbor parameters, including all possible mismatches in all possible trimer contexts (*17*). Sequence context is most likely important for PNA−DNA duplexes as well, even though the effects might be different. Here, although we do not intend to quantify the context dependence for either matched or mismatched PNA−DNA duplexes, we highlight a few instances where context dependence seems to be an important factor for interpretation of our data.

The simplest step in introducing context dependence is to account for the stability around the mismatch site, i.e., the distribution of A·T and G·C base pairs. For DNA−DNA duplexes, it has been observed that less stable nearest neighbors (A·T) can increase the negative influence of a mismatched base pair (*31*−*33*), and this seems to be valid for PNA−DNA systems as well. Inspecting Table 2, we indeed see, for example, that the A·G mismatch surrounded by two G·C base pairs costs less than the two others with one A·T and one G·C mismatch. Furthermore, the best agreement between our work and that of Kilså Jensen et al. (*5*) and Igloi (*19*) is found when we have surroundings with similar stabilities, e.g., for the C·X mismatches (Table 2).

The next step is to compare the detailed context of the different mismatches. In our P9/D9AX mismatch series, the neighbors on the DNA strand are T and G (AAC/TXG; see Figure 2b); in the work by Igloi (*19*), the surrounding base pairs of the mismatched position were A and G on the DNA strand, i.e., TAC/AXG, and in the work by Kilså Jensen et al. (*5*), the trimer context is CAC/GXG. Could these differences be the reason for the altered order of stabilities of the A·X mismatched PNA−DNA duplexes? For DNA−DNA systems, the stability ranking of mismatches is A·G > A·A > A·C for all three contexts (*17*), but considering the probable differences in structure on a detailed level, the effects are not unlikely to be different for PNA−DNA duplexes, thereby yielding a different order of stabilities in the mismatched A·X duplexes. For our mismatch series P9XT/D9 with the mismatched base on the PNA strand, we have the sequence AXC/TTG at the mismatch, which according to DNA−DNA nearest-neighbor data (*17*) would give decreasing stabilities in the series G·T > T·T > C·T, in good qualitative agreement in a relative sense with our observations. In fact, by analogy with DNA−DNA data (*17*), the G·T mismatch ought to be in its most favorable context in our sequence (GTT/CGA). As discussed previously, Igloi (*19*) found the somewhat different order T·T > G·T > C·T, but T_m data are taken from duplexes with the mismatch in three different contexts. Nevertheless, in all these contexts, DNA−DNA data (*17*) rank the G·T mismatch more stable than the T·T mismatch. While both mismatches have one A·T base pair and one G·C base pair around the mismatch, due to the inherent asymmetry of the PNA−DNA duplex base pair, comparison with DNA−DNA studies is difficult since no structural data are yet available for mismatches in PNA−DNA duplexes.

For our P12/D12CX series, the C·C and C·A mismatches are marginally more destabilizing than the C·T mismatch (Table 1). In this series, the ΔT_m pattern for mismatches in PNA−DNA duplexes is similar to that reported for DNA−DNA systems in the same context (ACA/TXT), i.e., C·C ≈ C·A and worse than C·T (*17*, *34*). Table 2 shows that our results on the C·X mismatches agree well with the work by Igloi (*19*) and Kilså Jensen et al. (*5*). The comparison is in this case more straightforward since all three results are for duplexes where the mismatch is surrounded by two A·T base pairs, even though our sequence contains the second most destabilizing context in DNA−DNA systems (*17*), ACA/TXT, while the two others have ACT/TXA, which is the most destabilizing context in DNA−DNA systems (*17*).

Thermodynamic Cost of a Single Mismatch. On the basis of the results obtained from melting curve analysis on different mismatches, we now discuss the specificity of binding in terms of a simple binding model. To estimate the binding constant of a PNA−DNA duplex, we factorize the total binding constant into individual (monomer) binding constants, according to the following convention where [△▽] represents the concentration of one fully matched base pair and [▽▽] denotes the concentration of one mismatched base pair. [▽] and [△] represent the concentration of free (non-base-pairing) monomers. Subscript mm denotes a mismatched base pair. The individual binding constants (lowercase letters used to distinguish them from the total binding constant) are defined as follows, for a matched and a mismatched base pair, respectively.

$$\triangle + \triangledown \overset{k}{\rightleftharpoons} \triangle\triangledown;\ k = \frac{[\triangle\triangledown]}{[\triangle][\triangledown]} \tag{4}$$

$$\triangledown + \triangledown \overset{k_{mm}}{\rightleftharpoons} \triangledown\triangledown;\ k_{mm} = \frac{[\triangledown\triangledown]}{[\triangledown][\triangledown]} \tag{5}$$

For the duplex, a factorized binding constant is obtained by taking the product of the individual binding constants of *N*

$-M$ matched base pairs, and M mismatched base pairs to generate a duplex of length N, according to eq 6, which shows the ratio (K/K_{mm}) of binding constants between the matched and mismatched duplex.

$$\frac{K}{K_{mm}} = \frac{k^N}{k^{N-M}k_{mm}{}^M} = \frac{k^{N-M}k^M}{k^{N-M}k_{mm}{}^M} = \frac{k^M}{k_{mm}{}^M} \qquad (6)$$

where the third ratio is given only to show how the final result is obtained. Consequently, from comparison of the binding constants of (singly) mismatched (K_{mm}) and matched (K) duplexes, and the average binding constant per matched base pair (k), an estimate of the binding constant of one mismatched base pair, k_{mm}, can be obtained.

For P9/D9AX, the binding free energy $\Delta G°$ of the mismatches is around 11 kJ mol^{-1} less favorable than for the perfectly sequence-matched duplex (Table 1). The decrease in binding constant K per mismatch (K/K_{mm}) converts to, on average, almost 2 orders of magnitude ($K/K_{mm} \approx 57$), which would correspond to ~1.5 matched average base pairs as concluded from the pure length dependence (Figure 3a). According to our simple binding model (eqs 4−6), this corresponds to an average binding constant of a single (A·X) mismatch (k_{mm}) of around 14 M^{-1}/57 (=0.24 M^{-1}). This is on the same order of magnitude as previously reported and lies between those found by Igloi (*19*), whose data led to an average k_{mm} of around 0.1 M^{-1}, and Kilså Jensen et al. (*5*), whose data result in an average k_{mm} of 0.3 M^{-1} (K/K_{mm} values given in Table 2). The higher values found using data from Kilså Jensen et al. (*5*) are most probably due to their A·X mismatch being surrounded by two G·C base pairs, whereas both we and Igloi (*19*) had one A·T and one G·C base pair neighboring the mismatch. Also, DNA−DNA nearest-neighbor data support this idea that a less stable mismatch neighborhood increases the thermodynamic cost of introducing a mismatch (*17*).

Thermodynamic analysis of the melting curves for the series P9XT/D9 indicated that the ratio K/K_{mm} (i.e., the cost of mismatch) increases quite considerably when C·T and T·T mismatches are introduced, whereas the G·T mismatch produces an only 10-fold increase (Table 2). From these values, k_{mm} is determined to be around 0.013, 0.02, and 1.3 M^{-1}, respectively. The value for C·T is in excellent agreement with that of Igloi (*19*) (k_{mm} = 0.015 M^{-1}), while the T·T mismatch differs considerably (k_{mm} = 0.49 M^{-1}) as does the ΔT_m (Table 2). Data from Kilså Jensen et al. (*5*) (Table 2) yield a k_{mm} of 0.051 M^{-1} for C·T, 4.3 M^{-1} for T·T, and 0.59 M^{-1} for G·T. Again, the T·T mismatch shows the largest deviation; however, our value for G·T is in good agreement with that of Kilså Jensen et al. (*5*), while thermodynamic data for the G·T mismatch are lacking in the study by Igloi (*19*). For the C·T mismatch, our result is closer to that of Igloi (*19*) most likely due to the neighboring sequence being two A·T base pairs in both cases. As already discussed, the correlation with mismatched DNA−DNA systems is quite extensive, indicating that the sequence context we have analyzed is extraordinarily stable for G·T mismatches (*17*), leading us to speculate that the conformational properties of the PNA backbone allow it to better accommodate this mismatch than a DNA backbone.

However, for our 12-mer P12/D12CT duplex, despite having two A·T base pairs surrounding the mismatch, the stability of the duplex decreases dramatically (see Table 2), giving a k_{mm} of 0.0008 M^{-1}, which is much lower than that of the C·T mismatch in our 9-mer system. This is most likely related to the total duplex stability (about 3 orders of magnitude higher than for the 9-mer), and it could also be pointed out that variations between mismatches are indeed smaller in this case.

Our results for the T·T mismatch, as discussed in the T_m section, exhibit the most striking difference upon comparison with other studies (Table 2). We obtain for the T·T mismatch a binding constant k_{mm} of 0.02 M^{-1}, a value around 2 orders of magnitude lower than in the works of Igloi (*19*) and Kilså Jensen et al. (*5*). We cannot explain this difference, other than in speculating that the smaller size of the T·T mismatched base pair decreases the duplex stability strongly, as can also be concluded from a PNA−PNA mismatch assay showing that the T·T mismatch was much more costly (7−10 °C) than both G·T and T·G for duplex stability (*7*), and indicated from T_m data for PNA−RNA duplexes (*5*) where the T·T mismatch was 1−3 °C less stable thermally than the G·T and T·G mismatches. We suggest that the PNA−DNA duplex containing a T·T mismatch would be an ideal candidate for a structural study investigating the basis of mismatch sensitivity in hybrid duplexes. The other pyrimidine·pyrimidine C·C mismatch induces an even larger decrease in the binding constant, giving a k_{mm} of 0.0002 M^{-1}, again a value about 2 orders of magnitude lower than those obtained by Igloi (*19*) and Kilså Jensen et al. (*5*), who found that k_{mm} = 0.01 and 0.02 M^{-1}, respectively. However, in this case, the neighboring base pairs are both A·T in all three cases, so probably even next-nearest neighbors affect the impact of the mismatch. A mismatch is likely to disturb the optimal stacking and hydrogen bonding patterns for more than the mismatched base pair. Mismatches containing C bases have been shown to be the most destabilizing in the DNA−DNA duplexes (*35*), and our data for the P9/D9CX series suggest the same to be true for the PNA−DNA case.

Enthalpy−Entropy Compensation. In Figure 6a, $\Delta H°$ values are plotted versus $T\Delta S°$ for three of the mismatched series (P9/D9AX, P9XT/D9, and P12/D12CX) and for all of the different lengths of matched duplexes (6−20-mer). All data points fit well to a line, indicative of enthalpy−entropy compensation. This is seen from parameters obtained from both UV melting curves and ITC experiments (data not shown) with P12/D12CX systems, even though the absolute values of the thermodynamic parameters can vary with the two techniques. Compensation between enthalpy (mainly hydrogen bonding energies and van der Waals interactions) and entropy (mostly rearrangements of the molecules, solvent water, and counterions) is observed for many different biological systems, e.g., drug−receptor binding (*36*) and DNA duplexes with exocyclic cytosine adducts (*37*) and for short-strand DNA oligomers with guanidinium-linked nucleosides (*38*). The interpretation of this behavior is normally expressed in terms of tighter binding (larger $-\Delta H°$) accompanied by a larger loss of degrees of freedom (larger $-\Delta S°$) in a compensatory manner, resulting in relatively small changes in binding free energy ($\Delta G°$). Considering only the two strands, a strand association process (dimerization) is per se entropically unfavorable, whereas the formation of hydrogen bonds (base pairs) between strands is enthalpically favored as is the stacking of bases. The

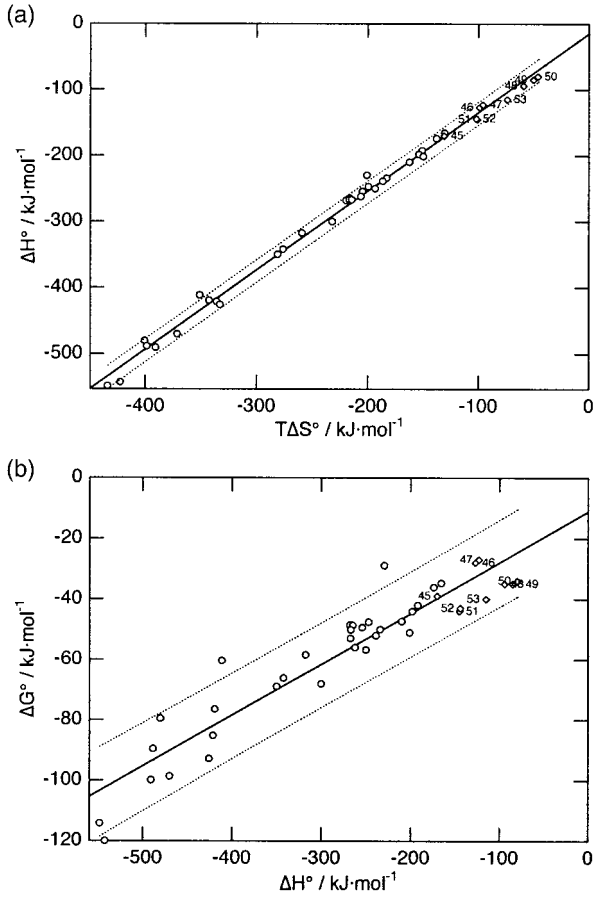

FIGURE 6: (a) Thermodynamic data ($\Delta H°$ vs $T\Delta S°$) showing the enthalpy−entropy compensation based on merging the data subset from Figure 3 (matched sequences 1−31, Figure 2a) and data for the mismatched 9- and 12-mer systems P9/D9AX, P9XT/D9, and P12/D12CX (points 45−52). The dotted curves show the 90% prediction interval (model plus random noise) for the $\Delta H°$ vs $T\Delta S°$ fit: $\Delta H° = (1.20 \pm 0.014)T\Delta S° - (15 \pm 3.5)$ kJ mol^{-1}. (B) Plot of the corresponding $\Delta G°$ vs $\Delta H°$ pairs. The dotted curves show the 90% prediction interval (model plus random noise) for the $\Delta G°$ vs $\Delta H°$ fit: $\Delta G° = (0.17 \pm 0.01)\Delta H° - (11 \pm 3)$ kJ mol^{-1}.

classical analysis of data is plotting $\Delta H°$ versus $T\Delta S°$, and if the data points correlate linearly, this is termed enthalpy−entropy compensation. However, as pointed out previously, a linear plot of the corresponding $\Delta H°$ versus $\Delta G°$ values is a preferable criterion in that the thermodynamical compensation has a chemical nature and is not due to statistical coincidence effects of the data acquisition (*36*, *39*, *40*). Our data suggest that the compensation is truly chemically based, relying on the fundamental nature of binding between the strands as base pairs are formed (see Figure 6b). Introducing a single mismatch significantly reduces the binding enthalpy as fewer hydrogen bonds are formed, and often also the stacking efficiency decreases. The loosening of the bonds (smaller $-\Delta H°$) at and in the vicinity of the mismatch is compensated by a less rigid structure (smaller $-\Delta S°$) upon forming a less perfect sequence-matched duplex.

When this simple picture is expanded upon, and as pointed out by Zhong et al. (*41*), most probably water structure (hydration) is altered significantly around the mismatched region of a DNA duplex, and also around the PNA−DNA duplexes studied here. These changes include immobilization of electrostricted water (enthalpically favored) and release of structural water (slightly unfavored enthalpically). At the same time, the order of the system increases, making an (compensating) unfavorable entropic contribution (*41*). Unfortunately, lacking reliable structural data, one can only speculate about the water structure (and the change thereof) around PNA−DNA duplexes containing mismatches. However, molecular dynamics simulations have indicated that the water structure around the somewhat more hydrophobic PNA backbone could be significantly different from that of DNA, i.e., in general energetically less favorable (*42*). In addition, it is likely that the water structure around the mismatched regions is different for the various mismatched base pairs. We observe quite different thermodynamic parameters for the mismatches (Table 1), but in the end, the unfavorable enthalpic contributions (loosening of intermolecular bonds) when introducing mismatches are still proportionally compensated by a favorable change in the degrees of freedom (the entropic factor). This is not surprising since most of the duplexes, mismatched or not, have at least six perfectly matched base pairs (counting both nearest neighbors surrounding the mismatched position) which all bind with the same kind of hydrogen bonding between bases and intrastrand stacking.

We have used a two-state model and assumed that $\Delta C_p = 0$ to determine the enthalpy and entropy changes associated with hybridization of PNA to DNA (see Materials and Methods). This is equivalent to assuming temperature-independent $\Delta H°$ and $\Delta S°$ values, which strictly speaking are extracted from the curve fits at T_m, and thus valid only at T_m when $\Delta C_p \neq 0$. This is generally an accepted procedure in nucleic acid work; however, there has recently been some concern about the validity of this assumption for DNA−DNA duplexes (*25*, *26*, *43*). We have followed the procedure of Rouzina et al. (*26*) to estimate ΔC_p from thermodynamic data obtained for the fully complementary PNA−DNA duplexes studied here. Hence, ΔC_p was estimated from plots of $\Delta H°$ versus T_m, and $\Delta S°$ versus ln T_m, and averaging the slopes in our case gives a $-\Delta C_p$ of 157 ± 36 J K^{-1} mol^{-1} bp^{-1}. This value appears to be smaller than even the lowest ones estimated for DNA−DNA hybridization. According to Rouzina et al. (*26*), the important factor to consider is the ratio $\Delta C_p/\Delta S°$, which with DNA−DNA data is 2−4; on the other hand, our data are in the low end of this range, giving a $\Delta C_p/\Delta S°$ ratio of $\approx$2.

Furthermore, attempts to determine ΔC_p from fitting the individual melting curves to a full analytical expression containing ΔC_p gave strongly varying results, often with low values of ΔC_p for the best fits, although T_m was practically unaffected. We have also performed single experiments with DSC on some decameric PNA−DNA duplexes (data not shown), without seeing any severe indications that there is a large ΔC_p involved. In a previously published DSC study on PNA−DNA and DNA−DNA duplexes, large variations in ΔC_p were obtained (*44*). For example, for PNA−DNA, $-\Delta C_p$ ranging from 0 to 652 J K^{-1} mol^{-1} bp^{-1} are reported, depending on concentration, which indicates that caution must be exercised when trying to determine ΔC_p values exactly.

Significant heat capacity effects are important if one wants to interpret the enthalpic and entropic contributions to the hybridization. However, because of enthalpy−entropy compensation, the individual contributions to $\Delta H°$ and $\Delta S°$ largely cancel out in $\Delta G°$. Therefore, for most applications

where only T_m (and/or $\Delta G°$) values are used, the simplest model with the historically accepted assumption that $\Delta C_p = 0$ will give sufficiently accurate analysis of PNA−DNA hybridization for comparative purposes.

CONCLUSIONS

The dependency of the free energy of hybridization of 31 mixed purine·pyrimidine sequence PNA−DNA duplexes of duplex length (6−20 bp) indicates an average binding constant per base pair of 14 M^{-1} which may serve as a guide for estimating duplex stability for perfectly matched duplexes.

The decrease in the macroscopic binding constant when introducing a single mismatch was found to vary depending on context: from a factor of 6×10^4 for a central C·C mismatch in a 12-mer duplex to a factor of just 10 for a central G·T mismatch in a 9-mer duplex. Thus, mismatch stability is highly context-dependent which has to be taken into account when designing PNA oligomers for targeting DNA or RNA sequences in both diagnostic and therapeutic applications.

On average, the apparent binding constant per mismatched base pair was as low as 0.02 M^{-1} which explains the high degree of sequence specificity generally observed for PNA binding to a single-stranded DNA target. The average cost of a single mismatch is therefore estimated to be on the order of, or larger than, the gain of two matched base pairs.

The composition of neighboring base pairs was found to be a primary determinant of mismatch stability where flanking G·C pairs seem in general to stabilize the mismatch relative to flanking A·T base pairs. In this respect, mismatches in PNA−DNA duplexes exhibit a behavior similar to that of DNA−DNA (*17*) and RNA−RNA duplexes (*45*).

Finally, the analysis of all determined thermodynamic parameters revealed that strong enthalpy−entropy compensation occurs in PNA−DNA duplexes. The weakening of bonds (smaller $-\Delta H°$) at and in the vicinity of the mismatch is compensated by a less rigid structure (smaller $-\Delta S°$) upon forming a less perfect sequence-matched duplex. This is likely to be an effect of not only changes in base pair formation and stacking but also rearrangements of solvent molecules and ions.

SUPPORTING INFORMATION AVAILABLE

Circular dichroism (CD) spectra (Figure S1) of the 9-mer mismatch series P9/D9AX and P9XT/D9 together with a short discussion of the results. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Nielsen, P. E. (1999) *Curr. Opin. Struct. Biol. 9*, 353−357.
2. Nielsen, P. E., and Egholm, M. (1999) in *Peptide Nucleic Acids (PNA). Protocols and Applications* (Nielsen, P. E., and Egholm, M., Eds.) Horizon Scientific, Norfolk, CT.
3. Nielsen, P. E., Egholm, M., Berg, R. H., and Buchardt, O. (1991) *Science 254*, 1497−1500.
4. Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B., and Nielsen, P. E. (1993) *Nature 365*, 566−568.
5. Jensen, K., Ørum, H., Nielsen, P. E., and Nordén, B. (1997) *Biochemistry 36*, 5072−5077.
6. Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P. E., Nordén, B., and Gräslund, A. (1996) *J. Am. Chem. Soc. 118*, 5544−5552.
7. Eriksson, M., and Nielsen, P. E. (1996) *Q. Rev. Biophys. 29*, 369−394.
8. Good, L., and Nielsen, P. E. (1997) *Antisense Nucleic Acid Drug Dev. 7*, 431−437.
9. Rufer, N., Dragowska, W., Thornbury, G., Roosnek, E., and Lansdorp, P. M. (1998) *Nat. Biotechnol. 16*, 743−747.
10. Weiler, J., Gausepohl, H., Hauser, N., Jensen, O. N., and Hoheisel, J. D. (1997) *Nucleic Acids Res. 25*, 2792−2799.
11. Wang, J., Rivas, G., Cai, X., Chicharro, M., Parrado, C., Dontha, N., Begleiter, A., Mowat, M., Palecek, E., and Nielsen, P. E. (1997) *Anal. Chim. Acta 344*, 111−118.
12. Ørum, H., Nielsen, P. E., Egholm, M., Berg, R. H., Buchardt, O., and Stanley, C. (1993) *Nucleic Acids Res. 21*, 5332−5336.
13. Veselkov, A. G., Demidov, V. V., Nielsen, P. E., and Kamenetskii, F. M. D. (1996) *Nucleic Acids Res. 24*, 2483−2487.
14. Seeger, C., Batz, H.-G., and Ørum, H. (1997) *BioTechniques 23*, 512−516.
15. Ratilainen, T., Holmén, A., Tuite, E., Haaima, G., Christensen, L., Nielsen, P. E., and Nordén, B. (1998) *Biochemistry 37*, 12331−12342.
16. Nielsen, P. E., and Christensen, L. (1996) *J. Am. Chem. Soc. 118*, 2287−2288.
17. Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J. (1999) *Biochemistry 38*, 3468−3477.
18. Giesen, U., Kleider, W., Berding, C., Geiger, A., Ørum, H., and Nielsen, P. E. (1998) *Nucleic Acids Res. 26*, 5004−5006.
19. Igloi, G. L. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 8562−8567.
20. Christensen, L., Fitzpatrick, R., Gildea, B., Petersen, K. H., Hansen, H. F., Koch, T., Egholm, M., Buchardt, O., Nielsen, P. E., Coull, J., and Berg, R. H. (1995) *J. Pept. Sci. 3*, 175−183.
21. Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M. (1986) in *Data for biochemical research*, pp 103−114, Oxford University Press, New York.
22. Nordén, B., Kubista, M., and Kurucsev, T. (1992) *Q. Rev. Biophys. 25*, 51−170.
23. Applequist, J., and Damle, V. (1965) *J. Am. Chem. Soc. 87*, 1450−1458.
24. Puglisi, J. D., and Tinoco, J. I. (1989) *Methods Enzymol. 180*, 304−325.
25. Rouzina, I., and Bloomfield, V. A. (1999) *Biophys. J. 77*, 3252−3255.
26. Rouzina, I., and Bloomfield, V. A. (1999) *Biophys. J. 77*, 3242−3251.
27. Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.
28. Draper, D. E., and Gluick, T. C. (1995) *Methods Enzymol. 259*, 281−351.
29. SantaLucia, J., Allawi, H. T., and Seneviratne, A. (1996) *Biochemistry 35*, 3555−3562.
30. Griffin, T. J., and Smith, L. M. (1998) *Anal. Biochem. 260*, 56−63.
31. Allawi, H. T., and SantaLucia, J. (1997) *Biochemistry 36*, 10581−10594.
32. Allawi, H. T., and SantaLucia, J. (1998) *Biochemistry 37*, 9435−9444.
33. Allawi, H. T., and SantaLucia, J. (1998) *Biochemistry 37*, 2170−2179.
34. Jackson, B. A., Alekseyev, V. Y., and Barton, J. K. (1999) *Biochemistry 38*, 4655−4662.
35. Aboul-ela, F., Koh, D., and Tinoco, I. J. (1985) *Nucleic Acids Res. 13*, 4811−4824.
36. Gilli, P., Ferretti, V., Gilli, G., and Borea, P. A. (1994) *J. Phys. Chem. 98*, 1515−1518.
37. Gelfand, C. A., Plum, G. E., Grollman, A. P., Johnson, F., and Breslauer, K. J. (1998) *Biochemistry 37*, 12507−12512.
38. Blasko, A., Dempcy, R. O., Minyat, E. E., and Bruice, T. C. (1996) *J. Am. Chem. Soc. 118*, 7892−7899.
39. Krug, R. R., Hunter, W. G., and Grieger, R. A. (1976) *J. Phys. Chem. 80*, 2341−2351.
40. Krug, R. R., Hunter, W. G., and Grieger, R. A. (1976) *J. Phys. Chem. 80*, 2335−2341.

41. Zhong, M., Marky, L. A., Kallenbach, N. R., and Kupke, D. W. (1997) *Biochemistry 36*, 2485−2491.
42. Sen, S., and Nilsson, L. (1998) *J. Am. Chem. Soc. 120*, 619−631.
43. Chalikian, T. V., Völker, J., Plum, G. E., and Breslauer, K. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 7853−7858.
44. Chakrabarti, M. C., and Schwarz, F. P. (1999) *Nucleic Acids Res. 27*, 4801−4806.
45. Zhu, J., and Wartell, R. M. (1997) *Biochemistry 36*, 15326−15335.

BI000039G