Sequence-Dependent B↔A Transition in DNA Evaluated with Dimeric and Trimeric Scales

M. Y. Tolstorukov,* V. I. Ivanov,† G. G. Malenkov,‡ R. L. Jernigan,* and V. B. Zhurkin*

*Laboratory of Computational and Experimental Biology, National Cancer Institute, NIH, Bethesda, Maryland 20892 USA;
†W. A. Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991 Russia; and
‡Institute of Physical Chemistry, Russian Academy of Sciences, Moscow, 119991 Russia

ABSTRACT Experimental data on the sequence-dependent B↔A conformational transition in 24 oligo- and polymeric duplexes yield optimal dimeric and trimeric scales for this transition. The 10 sequence dimers and the 32 trimers of the DNA duplex were characterized by the free energy differences between the B and A forms in water solution. In general, the trimeric scale describes the sequence-dependent DNA conformational propensities more accurately than the dimeric scale, which is likely related to the trimeric model accounting for the two interfaces between adjacent base pairs on both sides (rather than only one interface in the dimeric model). The exceptional preference of the B form for the AA:TT dimers and AAN:N′TT trimers is consistent with the cooperative interactions in both grooves. In the minor groove, this is the hydration spine that stabilizes adenine runs in B form. In the major groove, these are hydrophobic interactions between the thymine methyls and the sugar methylene groups from the preceding nucleotides, occurring in B form. This interpretation is in accord with the key role played by hydration in the B↔A transition in DNA. Importantly, our trimeric scale is consistent with the relative occurrences of the DNA trimers in A form in protein-DNA cocrystals. Thus, we suggest that the B/A scales developed here can be used for analyzing genome sequences in search for A-philic motifs, putatively operative in the protein-DNA recognition.

INTRODUCTION

DNA conformational polymorphism was detected at the early stage of double helix history (Franklin and Gosling, 1953). Depending on the relative humidity (RH) of the environment, the DNA in fibers and films is stabilized either in A form (low RH) or in B form (high RH) (Franklin and Gosling, 1953; Lindsay et al., 1988; Maleev et al., 1993). Decreasing RH by adding alcohol or some salts to the DNA solution also induces the B→A transition (Ivanov et al., 1974; Nishimura et al., 1986). Comparison of this transition for various water-alcohol mixtures has led to the concept that the water activity (or chemical potential) is an apparent driving force of the B↔A transition (Malenkov et al., 1975). This reflects the important role of hydration in stabilizing a specific DNA conformation.

The transition between the A and B forms of DNA is reversible, cooperative, and sequence specific (for review, see Ivanov and Minchenkova, 1995). In particular, those authors demonstrated that poly(dG)·poly(dC) undergoes the B→A transition most easily (in their words, this polymer is "A-philic"), whereas poly(dA)·poly(dT) resists the B→A transition (this polymer is "B-philic"). This finding is biologically relevant, as the propensity of a given nucleotide sequence toward A or B conformation is one of the key factors in the protein-DNA interactions (Lu et al., 2000; Olson and Zhurkin, 2000). In particular, when cyclic AMP receptor protein (CRP) binds to a sequence with a long spacer between the "consensus boxes," the DNA undergoes a B-to-A-like transition (Ivanov et al., 1995). Importantly, this binding is stronger when the central part of the DNA is A-philic.

The functional importance of the B→A transition was amply demonstrated when it was found that the growing DNA duplex in the active centers of DNA polymerase (Kiefer et al., 1998) and HIV-1 reverse transcriptase (Ding et al., 1998) is in the A form. Besides, the RNA-DNA heteroduplex transiently formed during transcription is also in the A form (Cheetham and Steitz, 1999). Stabilization of the A form is believed to be critical for increasing the fidelity of DNA and RNA synthesis (Timsit, 1999).

Various models have been proposed to describe the sequence-dependence of the B↔A transition in solution. Ivanov and co-workers have extensively used the dimeric model, assuming that the B↔A transition sequence specificity is determined only by contacts between adjacent nucleotide pairs (Ivanov and Minchenkova, 1995, and references therein). Recently, a more complicated trimeric code for the A-propensity was suggested based on calculations of the accessible surface areas for various trimers (Basham et al., 1995). Although the latter model has certain advantages related to the inclusion of effects from two neighbors to a pair instead of one, it has not been adequately tested against the available experimental data.

In the present study we seek rules for B/A-philicity deduced from experimental data on the alcohol-induced B→A transition in oligomeric DNA duplexes within the framework of dimeric and trimeric models. The goal is a systematic compar-

Received for publication 26 April 2001 and in final form 18 September 2001.

Address reprint requests to V. B. Zhurkin, Laboratory of Computational and Experimental Biology, National Cancer Institute, NIH, Building 12B, Room B116, 12 South Drive, Bethesda, MD 20892; Tel.: 301-496-8913; Fax: 301-402-4724; E-mail: zhurkin@nih.gov.

0006-3495/01/12/3409/13 $2.00

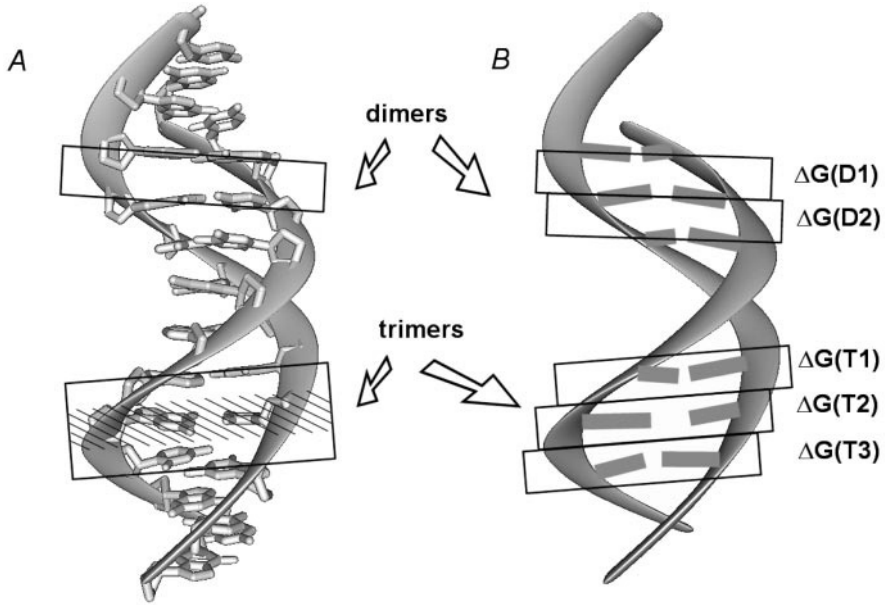

FIGURE 1 Dimeric and trimeric "steps" used to model the sequence dependence of the B↔A transition. The Drew-Dickerson dodecamer is shown: (*A*) ribbon representation with distorted base pairs and sugar rings visible; (*B*) schematic view emphasizing nonuniform buckles in the base pairs. Dimeric steps (*A* and *B*, above) account only for the contacts between consecutive nucleotide pairs, e.g., for the stacking interactions between bases. The consecutive dimers (*D1* and *D2*) do not overlap, so their free energies, $\Delta G(D1)$ and $\Delta G(D2)$ are additive (*B*, above). A trimeric step includes both the contacts between pairs and the nucleotide pair itself, that is, the base pair distortions, sugar puckering, etc., are implicitly incorporated in the model (*shaded* in *A*, below). The consecutive trimers overlap, therefore, to make their free energies additive, the energies of contacts are divided between two adjacent trimers. To illustrate this point, the trimers are shown without overlaps (*B*, below). Then, the free energy of the oligomer can be decomposed into the sum of free energies of the trimers: $\Delta G(T1)$, $\Delta G(T2)$, $\Delta G(T3)$, etc.

ison of dimeric and trimeric models against the same set of experimental data and theoretical formalism.

MATERIALS AND METHODS

Model and the experimental dataset

A simple way to account for the sequence specificity of the B↔A transition is to use a phenomenological approach based on the one-dimensional Ising model. In this case, a DNA duplex is decomposed into consecutive "sequence steps," such that the energy of transition for the whole duplex could be represented as a linear combination of the "step" energies along the molecule. The "sequence steps" considered in the present study are either dimers or trimers, that is, two or three consecutive nucleotide pairs (Minchenkova et al., 1986; Basham et al., 1995) (Fig. 1). Decomposition of a duplex into its monomeric units (as it was done in early helix-coil models) is based on the implicit assumption that all pairs interact with each other with the same energy, i.e., the internucleotide contacts are sequence independent. Notice, however, that due to the difference in the stacking interactions in B- and A-DNA, it is necessary to take into account the sequence specificity of these contacts for modeling the B↔A transition (Minchenkova et. al., 1986; Ivanov and Minchenkova, 1995). Therefore, the "steps" considered here contain more than one nucleotide pair.

According to the model considered, the free energy of the B↔A transition in a duplex can be represented as:

$$\Delta G_{\mathrm{BA}}^{\mathrm{theor}} = \sum_{I \in \text{all steps}} \Delta G_{\mathrm{BA}}(I) + 2\Delta G^{2}, \qquad (1)$$

in which $\Delta G_{\mathrm{BA}}(I)$ is the energy change of the I-th "sequence step" (dimer or trimer), and $2\Delta G^{s} = 3.0$ kcal/mol is the free energy change due to contacts of the two terminal pairs with solvent (Minchenkova et al., 1986). For simplicity, the latter term is assumed to be sequence nonspecific. This formula is applicable if the duplex is shorter than the cooperativity length of the B↔A transition, 15 to 20 basepairs. (Minchenkova et al., 1986), in other words, when the transition occurs in an "all-or-nothing" way. This equation can be also used for long polymers with "simple" repeating sequences, such as poly(dA-dC)·poly(dG-dT); in this case, the second term has to be omitted.

Experimentally, the free energy change can be estimated from the data on the B→A transition in solution, as forced by an increase in alcohol concentration. To this aim, an environmental parameter, a, has to be selected such that the free energy changes characterizing the B↔A transition, ΔG_{BA}, would depend linearly on a. It was shown that for ethanol and some other alcohols, the water activity can be used as such a "linearizing parameter" (Malenkov et al., 1975). Then, the energy values can be represented in the following way (Minchenkova et al., 1986). At the midpoint of the transition, ΔG_{BA} equals zero, thus for any duplex:

$$\Delta G_{\mathrm{BA}}^{\mathrm{est}} = \frac{RT}{Q}\, n(a - a_{\mathrm{t}}), \qquad (2)$$

in which a_{t} is the value of the parameter a at the midpoint of the B↔A transition, Q is an empirically determined coefficient, R is the gas constant, and $T = 253$ K is the experimental temperature (kept below zero to prevent melting of short oligomers). The value n is comprised of the "real" sequence steps in the duplex and the two tails of an oligomer. Thus, for a trimeric model, n equals the number of base pairs and the number of base pairs plus one for a dimeric model.

At present, trifluoroethanol (TFE) is being used to effect the B→A transition, because in this case it is easier to avoid DNA precipitation compared with other alcohols, such as ethanol. The sequence-specific effects of TFE and ethanol on the B↔A transition are similar, i.e., the A-philicity scales are

TABLE 1 Experimental B↔A transition midpoints and comparison with the dimeric and trimeric models

			Fitting deviations (% AWA)			
	Sequence*	Apparent water activity (% AWA)†	Dimeric model D-95‡	Dimeric model D-10‡	Dimeric model D-12‡	Trimeric model T-32‡
1	CCGATATCGG	75.2	0.059	−1.395	−0.823	0.000
2	CCAGATCTGG	73.8	−0.559	−1.864	−1.786	0.001
3	CCCTGCAGGG	78.7	0.559	0.080	0.165	0.001
4	GGGGGCCCCC	83.5	−0.265	−0.536	−0.659	0.002
5	TATCACCGC	76.0	2.484	−1.350	−0.062	0.001
6	GGACCCGGGTCC	82.7	0.319	0.351	0.285	0.000
7	CCAACGTTGG	75.0	−0.206	−0.853	−0.352	0.000
8	CCAAGCTTGG	73.5	0.147	−1.003	−0.996	0.000
9	CCGAATTCGG	73.0	−0.147	−1.147	−1.169	0.000
10	ACACCGGTGT	80.0	−0.147	0.320	0.612	0.011
11	GCTACGTAGC	77.2	−0.206	−1.655	−0.106	0.002
12	GGCAGCTGCC	78.0	3.206	0.013	0.941	−0.001
13	AGAGGCCTCT	77.4	−0.147	0.649	−0.518	0.000
14	ACTACTAGTAGT	80.0	−2.255	0.340	0.320	−0.009
15	ACTACTAGTACT	80.0	−2.255	0.340	0.320	0.010
16	CCCCCGGGGG	84.2	−0.324	1.255	0.804	−0.002
17	ATACCGGTAT	79.3	0.088	−0.016	0.277	−0.009
18	ACGCACCACG	77.6	−0.029	−0.501	0.449	−0.001
19	TGATGATCATCA	75.0	7.034	0.983	1.228	−0.001
20	ACTACCCGAAATGA	78.5	1.660	1.270	0.735	0.000
21	poly d(A-T)	81.5	0.735	0.001	0.000	0.044
22	poly d(G-C)	81.2	−0.882	0.004	−0.001	0.001
23	poly d(A-C)	82.5	−1.324	0.000	−0.004	−0.107
24	- TTTTTTG	74.1	2.164	0.793	0.716	0.000

Dimeric model D-95 was proposed by Ivanov and Minchenkova (1995).
D-10 and D-12 are the 10- and 12-parameter dimeric models developed here. T-32 is the 32-parameter trimeric model.
*Only one strand of the duplexes is given. In the last line, data for the B-philic fragment of the 14-mer d(AC_6T_6G):d(CA_6G_6T) are presented (Ivanov et al., 1996).
†Experimentally measured B↔A transition midpoints in % of apparent water activity (AWA). Experimental error is 0.5% AWA (Minchenkova et al., 1986).
‡Fitting deviations between the experimental and the computed values for the B↔A transition midpoints (in % AWA). Note that only trimeric model T-32 gives deviations smaller than the experimental error.

apparently close for these alcohols, at least for those sequences that were tested in both solutions (Ivanov et al., 1979, 1983). Because the water activity is unknown for the TFE solutions at low temperature (Cooney and Morcom, 1988), it is necessary to select independently an appropriate "linearizing parameter" a for TFE to be able to use Eq. 2 in this case. In particular, this selection has to guarantee that the coefficient Q does not depend on the DNA sequence. A careful analysis demonstrated that this requirement holds for decamers with various GC content (Minchenkova et al., 1986; Ivanov and Krylov, 1992), if one assumes that the parameter a as a function of the TFE concentration is close to the water activity in the water-ethanol mixture with the same concentration (Malenkov and Minasyan, 1977). This empirical parameterization yields the value of $Q = 14.5\%$ for $T = 253$ K (Eq. 2). Below, we refer to the obtained function a as apparent water activity (AWA) (Table 1).

Importantly, this thermodynamic function accounts not only for the activity (or chemical potential) of water in the TFE-water mixture but also for the relative preference for the B form versus A form in this solution, which depends on interactions of DNA both with water and alcohol. In other words, this empirical function can be considered as a "thermodynamic force" driving the B↔A transition.

Equating the theoretical and the experiment-based free energy changes, ΔG_{BA}^{theor} from Eq. 1 and ΔG_{BA}^{est} from Eq. 2, we obtain:

$$\sum_{I \in \text{all steps}} \Delta G_{BA}(I) + 2\Delta G^{s} = \frac{RT}{Q}\, n(a - a_t) \qquad (3)$$

Thus, having measured the transition points (a_t) for duplexes with different sequences, one obtains the set of linear equations to determine the free energy changes for all possible sequence steps, $\Delta G_{BA}(I)$.

In this study, we use data on the midpoints of the B↔A transition in water solution (induced by the change in alcohol content) for the set of 20 DNA oligomers and three polymers obtained as described by Minchenkova et al. (1986). Partially, these data have been published earlier (Minchenkova et al., 1986; Ivanov et al., 1983, 1985). Here, the whole dataset is presented for the first time (Table 1). Each of these oligomers is shorter than 15 basepairs and undergoes the B↔A transition as a single entity. Thus, the formalism of Eq. 1 is applicable.

In addition, we used the data for the 14-mer d(AC_6T_6G), which demonstrated a two-stage B→A transition (Ivanov et al., 1996). In this case, the transition midpoint for the B-philic moiety (5′-T_6G-3′) was considered. Eq. 1 was modified to account for the B/A junction at the 5′ end, and for one terminal pair at the 3′ end of this 7-mer. The B/A junction energy was 1.4 kcal/mol (Ivanov et al., 1996). Overall, our modeling was based on 24 measured B↔A transition midpoints (Table 1).

Minimization procedure

To resolve the set of 24 linear equations, we used the following minimization procedure

$$\Delta^{j} = \frac{RT}{Q}\, n^{j}(a - a_t^{j}) - \sum_{\substack{I \in \text{all steps} \\ \text{of j-th duplex}}} \Delta G_{BA}(I) + 2\Delta G^{s}$$

$$\sum_{j=1}^{24} (\Delta^{j})^2 \rightarrow \min \qquad (4)$$

TABLE 2 Free energy parameters for the dimeric models

Dimeric step	Number of occurrences*	ΔG_{BA} (kcal/mol) D-95†	D-10‡	D-10/12§	D-12§
AA:TT	13	**0.97**	**1.09** (0.04)	**0.98**	**1.39** (0.09)
GG:CC	48	**0.19**	**0.32** (0.01)	**0.22**	**0.89** (0.05)
AC:GT	27	**0.13**	**0.37** (0.14)	**0.35**	**0.89** (0.08)
CA:TG	22	**1.04**	**0.82** (0.14)	**0.84**	**1.38** (0.13)
AT:AT	12	**0.58**	**0.86** (0.31)	**0.73**	**1.14** (0.11)
TA:TA	14	**0.73**	**0.40** (0.31)	**0.53**	**0.94** (0.26)
AG:CT	23	**0.33**	**0.83** (0.14)	**0.65**	**1.19** (0.05)
GA:TC	17	**0.98**	**0.66** (0.17)	**0.65**	**1.19** (0.09)
CG:CG	15	**0.52**	**0.84** (0.14)	**0.78**	**1.45** (0.06)
GC:GC	12	**0.73**	**0.44** (0.14)	**0.50**	**1.16** (0.12)
ΔG^s		**1.625**	**1.50**	**1.50**	**1.50**
A:T		—	—	—	**−0.41** (0.10)
G:C		—	—	—	**−0.67** (0.06)
Average		**0.62**	**0.66**	**0.62**	

Values of ΔG_{BA} for the four dimeric models are given in bold, with their r.m.s.d. in parentheses.
*Number of occurrences of a given dimeric step in the experimental dataset.
†D-95 is the dimeric model suggested by Ivanov and Minchenkova (1995), based on the limited experimental dataset.
‡D-10 is the 10-parameter model obtained in the course of fitting the entire dataset available (see Table 1, Eq. 4).
§D-12 and D-10/12 are the 12-parameter models derived from Eqs. 5 and 6 respectively (see Results, Dimeric models).

The first expression defines the deviation between the experimental transition midpoint and the theoretical prediction for the *j*-th duplex (Eq. 3). The second expression defines the cumulative square error, that is, the sum of the square deviations for all 24 sequences. For both the dimeric and trimeric models the same set of experimental data was used (Table 1). There is no occurrence of the TAA:TTA trimer in the experimental data, so it was excluded from the fitting procedure (see Table 3). In the case of the dimeric model there are 10 variables $\Delta G_{BA}(I)$, corresponding to 10 unique dimers (Table 2). Thus, the set of 24 equations is over determined, and they can be solved only approximately. By contrast, in the case of trimeric model there are 32 possible trimers (Table 3), and the set of equations is under determined. As a consequence, Eqs. 3 and 4 have multiple solutions.

Initial conditions for minimization were chosen as the random uniform deviates in the interval 0.0 to 1.0. For each of 1000 random sets of initial conditions, three minimization cycles were carried out. Each cycle included 50,000 iterations (or it was interrupted if the cumulative square error changed by less than 10^{-5} in three successive iterations). Then, the obtained "best fits" were averaged (Tables 2 and 3). For the trimeric model only those sets of solutions were taken for averaging that gave cumulative error less than 0.5 kcal/mol, provided that all the individual deviations were less than 0.15 kcal/mol for oligomers and less than 0.03 kcal/mol for polymers. These values correspond to the differences in water activity less than 0.5%, which is the experimental error in measuring the B↔A transition midpoints (Minchenkova et al., 1986).

No such restraints were imposed on the "approximate solutions" of Eq. 3 for the dimeric model, because in the latter case the system of Eq. 3 is over determined, and the best fits are characterized by relatively large errors (Table 1).

RESULTS

Dimeric models

The ΔG_{BA} values obtained from the optimization procedure (Eq. 4) are presented in Table 2 (see model D-10). Importantly, the GG and AC dimers are the most A-philic, and the AA and CA are among the most B-philic dimers, entirely consistent with the model suggested earlier (Ivanov and Minchenkova, 1995), referred to here as the D-95 model. However, the other dimers behave differently in the two models (Fig. 2). According to our results, the AG, AT, and CG dimers are essentially B-philic ($\Delta G_{BA} > 0.8$); whereas in the early D-95 model these dimers were either intermediate or A-philic ($\Delta G_{BA} \leq 0.6$). By contrast, the dimers TA and GC were B-philic in the D-95 model ($\Delta G_{BA} = 0.73$), but they are intermediate in our model. In terms of graphic representation, the AG, CG, and AT dimers are positioned below the diagonal in Fig. 2, whereas TA and GC are above the diagonal. The likely reason for this inconsistency is that a larger set of experimental data are used here compared with the study by Ivanov and Minchenkova (1995).

Both dimeric models, D-95 and D-10, fail to fit all the data within the limit of experimental error, 0.5% of water activity. An especially large deviation, 7%, is observed for duplex 19 when the model D-95 is used (Table 1).

To improve the correspondence between theory and experiment, the dimeric model has been extended by introducing two additional parameters, ΔG_{BA}(A:T) and ΔG_{BA}(G:C), accounting for the free energy changes in the A:T and G:C pairs upon the B↔A transition (Table 2, model D-12). These parameters reflect the changes in nucleotide pairs themselves, involving such structural variables as sugar puckering, base pair buckle, etc. Then, Eq. 1 for the extended dimeric model is rewritten as:

$$\Delta G_{BA}^{theor} = \sum_{I \in \text{all dimeric steps}} \Delta G_{BA}(I) + 2\Delta G^s + m\Delta G_{BA}(\text{A:T}) + k\Delta G_{BA}(\text{G:C}), \quad (5)$$

in which m and k are the numbers of A:T and G:C pairs, respectively. Using this model, one gets a better fit to the

TABLE 3 Free energy parameters for the trimeric model

Trimeric step	Number of occurrences*	ΔG_{BA} kcal/mol
B-philic		
AAA:TTT	5	1.17 (0.08)
AAC:GTT	2	1.22 (0.34)
AAG:CTT	2	1.57 (0.38)
AAT:ATT	3	1.39 (0.35)
ATC:GAT	9	1.11 (0.20)
Intermediate		
ACA:TGT	3	0.52 (0.14)
ACG:CGT	6	0.57 (0.16)
AGA:TCT	4	0.95 (0.29)
AGC:GCT	6	0.48 (0.16)
AGG:CCT	4	0.64 (0.17)
ATA:TAT	7	0.63 (0.00)
CAA:TTG	5	0.52 (0.31)
CAC:GTG	6	0.67 (0.15)
CAG:CTG	6	0.46 (0.24)
CCA:TGG	7	0.44 (0.26)
CCG:CGG	14	0.42 (0.14)
CGA:TCG	5	0.56 (0.14)
CGC:GCG	4	0.64 (0.00)
CTA:TAG	10	0.66 (0.04)
GAA:TTC	3	0.72 (0.35)
GAC:GTC	2	0.49 (0.32)
GCA:TGC	5	0.76 (0.23)
GCC:GGC	6	0.54 (0.14)
GGA:TCC	2	0.56 (0.32)
GTA:TAC	10	0.67 (0.03)
TAA:TTA	0	N/A
TCA:TGA	6	0.58 (0.28)
A-philic		
ACC:GGT	9	0.30 (0.14)
ACT:AGT	9	0.30 (0.05)
ATG:CAT	3	0.22 (0.28)
CCC:GGG	17	0.26 (0.05)
CTC:GAG	2	0.21 (0.37)
ΔG^s		1.50
Average		0.65 (80.8)

The ΔG_{BA} values are averaged over the best fits obtained during minimization, Eq. 4. The r.m.s.d. are given in parentheses. Trimers are divided into three groups: B-philic ($\Delta G_{BA} > 1.1$), A-philic ($\Delta G_{BA} \leq 0.3$), and intermediate (all the rest). For graphic representation of the trimeric model see Fig. 1.

*Number of occurrences of a given trimeric step in the experimental dataset.

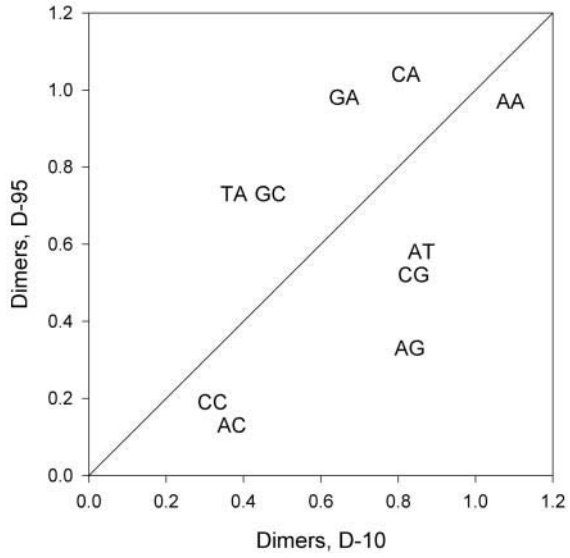

FIGURE 2 Comparison between two dimeric models: the D-95 model proposed by Ivanov and Minchenkova (1995), and the D-10 one obtained here by fitting the experimental data with Eq. 4 (Table 2). The ΔG_{BA} values are presented as scatter plot (kcal/mol). Here and in Fig. 3 the centers of two-letter labels are placed at the positions of the data points.

experimental data than for the models D-95 and D-10 (Table 1). For example, the number of duplexes for which the deviation between the computed and observed midpoints of the B↔A transition exceeds 1% of water activity and decreases from seven (for D-95) or eight (for D-10) to three for the extended model D-12.

At the same time, the B-philicity scales in the models D-12 and D-10 are practically the same. To illustrate this point, additional calculations are necessary, because the free energy parameters in the models D-10 and D-12 cannot be compared directly because they have different meanings. It is easy, however, to rearrange the energy terms in the D-12 model to obtain free energy values comparable with the ΔG_{BA} in the D-10 model. One has only to add the corresponding A:T and G:C pair free energy changes to the "contact" energies:

$$\Delta G_{BA}^{D-10/12}(MN) = \Delta G_{BA}^{D-12}(MN) + \frac{1}{2}[\Delta G_{BA}^{D-12}(M{:}M') + \Delta G_{BA}^{D-12}(N{:}N')]. \quad (6)$$

Here and below M and N stand for arbitrary nucleotides, and M′ and N′ for their complements. The ΔG values are presented in Table 2 (see model D-10/12).

The main result obtained earlier, that is, an extreme B-philicity for the AA:TT dimer and A-philicity for the GG:CC dimer, remains valid (Table 2). In addition, both dimeric models, the D-10 and the extended D-10/12, predict the AT dimer to be highly B-philic ($\Delta G_{BA}(AT) = 0.73$–0.86 kcal/mol), and the TA dimer to be intermediate ($\Delta G_{BA}(TA) = 0.40$–0.53 kcal/mol) (see Discussion below). The two models produced similar results for the other steps as well: the difference in $\Delta G_{BA}(MN)$ value is never more than 0.13 kcal/mol. Transition free energies in dimeric models D-10, D-10/12, and D-95 for all steps are positive, thus indicating that all sequences prefer B form in water solution.

Note that the G:C pair is more A-philic than A:T ($\Delta G_{BA}(G{:}C) < \Delta G_{BA}(A{:}T)$; Table 2; model D-12). This may be connected to the relative ease with which cytosine switches its conformation from C2′-endo to C3′-endo, as has been predicted by computations (Gorin et al., 1990;

Foloppe and MacKerell, 1999) and recently observed in the nuclear magnetic resonance study of 11-mer duplex (Kamath et al., 2000).

In conclusion, the dimeric models considered here are generally consistent with each other in identifying the most A- and B-philic dimers. Nevertheless, they still fail to predict the B↔A transition midpoints with accuracy better or equal to the experimental error (Table 1). Thus, a more detailed model is necessary.

Trimeric model

The free energy parameters for the trimeric model are presented in Table 3. All trimers are arbitrarily divided into three subsets: the B-philic trimers with $\Delta G_{BA} > 1.1$ kcal/mol, the A-philic ones with $\Delta G_{BA} \leq 0.3$ kcal/mol, and the "intermediate" trimers with the ΔG_{BA} values in between. Remarkably, the B-philic trimers are AT rich and many (but not all) contain the AA:TT dimer, one of the most B-philic dimers (Table 2). Furthermore, all the AAN:N′TT trimers are B-philic in our classification. Possible stereochemical reasons for this tendency are given in the following sections.

Notably, there are no GG:CC steps in the B-philic trimers (Table 3). Again, this is consistent with the dimeric model predicting extreme A-philicity for GG:CC. By contrast, this dimer occurs in two of the five most A-philic trimers, GGG:CCC and ACC:GGT. Overall, the A-philic trimers are GC rich (9/15 = 60%), whereas the B-philic trimers are AT rich (12/15 = 80%).

Deviations between the calculated free energy values and experimental data are all below the experimental error, 0.5% (Table 1). At the same time, the root mean square deviations (r.m.s.d.) are relatively large for all ΔG_{BA} values. This is apparently caused by the fact that the set of 24 equations (Eq. 3) is under-determined for the 32-parameter trimeric model, thus it has a continuum of solutions. Notably, the most B-philic trimers, AAN:N′TT, are among those with maximal r.m.s.d. (up to 0.38 kcal/mol; Table 3).

It is well known that DNA undergoes a partial B→A transition in complexes with proteins (Nekludova and Pabo, 1994; Shakked et al., 1994; Lu et al., 2000). Therefore, we decided to compare the predictions of the trimeric model with the conformation preferences of various sequences revealed in the protein-DNA cocrystals. To distinguish between the B and A forms of DNA, we took advantage of the structure-dependent parameter Z_p effectively separating the two DNA forms (El Hassan and Calladine, 1997; Lu et al., 2000). This parameter is defined as the z displacement of the phosphorus atom from the xy plane of the "middle" coordinate frame between neighboring base pairs.

Our analysis is based on the data presented by Lu et al. (2000). Among the protein and drug-bound DNA crystal complexes, they have selected those containing A-like conformational motifs; there were 42 structures overall made up of 39 proteins and three drugs. The Z_p value was calculated for each of the DNA dimeric steps (see Table 2 in Lu et al., 2000). To characterize the trimeric step conformations in the context of the B→A transition, we use the following criterion: a trimer is in the A form if both dimeric steps in this trimer have $Z_p > 0.5$ Å and at least one dimer has Z_p greater than 1.5 Å. In terms of Lu et al. (2000) this means that either both dimers are in the A form or one is in the A form and the other in the "intermediate transition zone."

For the three groups of trimers (B-philic, intermediate, and A-philic (Table 3)), the relative occurrence of the A form in these 42 complexes increases from 6 to 28 to 41%. Thus, the occurrence of the A form in the cocrystals strongly correlates with the A-philicity scale suggested here. In particular, the most A-philic trimer, CTC:GAG, is found in the A conformation in 71% of all cases, and this is the largest percentage among all 32 trimers. Using sugar pucker as a criterion for distinguishing the B and A forms (C2′-endo versus C3′-endo) gives a similar tendency (data not shown).

Importantly, the protein-DNA cocrystals were grown in the presence of glycerol and/or polyethylene glycol (PEG); the typical concentration of the latter was 15 to 25%. These organic compounds decrease the water activity and thus may facilitate the B→A transition, although the molecular mechanisms of their influence on the DNA conformation is probably different from those for TFE as considered here. Comparing the two sets of data obtained under different conditions, we conclude that both in solution (containing TFE) and in the protein-DNA cocrystals (containing PEG) the so-called A-philic DNA trimers (Table 3) undergo the B→A transition much more easily than do the B-philic trimers.

Comparison between the dimeric and trimeric models

To compare the two types of models, we "reduce" the trimeric code to the dimeric one by means of averaging:

$$\Delta G_{BA}^{D-10/32}(MN) = \frac{1}{8}\left[\sum_{K=A,T,G,C} \Delta G_{BA}^{T\text{-}32}(KMN) + \sum_{L=A,T,G,C} \Delta G_{BA}^{T\text{-}32}(MNL)\right]. \quad (7)$$

This procedure implies that the free energy change for the dimer MN, $\Delta G_{BA}(MN)$, should be equal to the average of the corresponding values for all the trimeric steps containing this dimer, denoted above $\boldsymbol{\Delta} G_{BA}^{D-10/32}$ (MN). In principle, this assumption is correct for an "infinitely large" set of oligomers with "random" sequences. In our case with 24 oligomers, however, the values ΔG_{BA} (MN) and $\Delta G_{BA}^{D-10/32}(MN)$ can only be approximately equal because

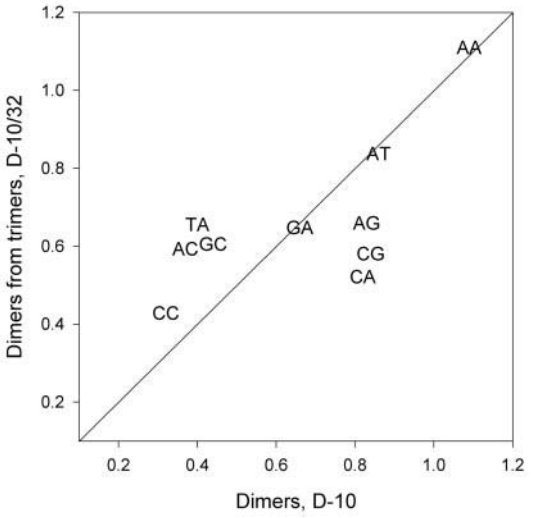

FIGURE 3 Comparison between the ΔG_{BA} values (in kcal/mol) for the dimeric model D-10 and those recalculated from the trimeric model using Eq. 7 (D-10/32).

of the nonuniform representation of the dimeric and trimeric steps in the sequences (Tables 2 and 3).

A comparison of the values ΔG_{BA} from the model D-10 with those recalculated from the trimers according to Eq. 7 is presented in Fig. 3. The major difference between the two sets is that the D-10/32 values have a narrower distribution. For example, the dimers AC, GC, CA, and CG all have $\Delta G_{BA}^{D-10/32} = 0.52$ to 0.61 kcal/mol, whereas in the D-10 model the corresponding values span the interval 0.37 to 0.84 kcal/mol. In other words, in the "reduced" D-10/32 model the sequence specificity of the B↔A transition is partially lost. All dimers are naturally divided into three sets: B-philic AA:TT (and perhaps, AT:AT), A-philic CC:GG, whereas the rest dimers are intermediate. In this respect, the D-10/32 model is similar to a simplified three-parameter dimeric model used by Ivanov and Krylov (1992) to simulate the B→A transition in the sea urchin 5S rRNA gene fragment observed by Becker and Wang (1988).

Overall, comparing various models, we see that the deviations between the experimental data and the values calculated using the trimeric model are much smaller than with the dimeric models (Table 1). The advantage of the trimeric model is evident from Fig. 4: compare the scatter for the dimeric models D-95 and D-10 with the straight line for the trimeric T-32.

Generally speaking, this advantage is a consequence of a greater number of the adjustable parameters in the trimeric model, allowing a more precise fitting of the experimental data. In terms of stereochemistry, this can be interpreted in the following way. Consider two trimers, CGC and CGT in B form (Fig. 5). The conformation of the central guanine depends on the neighbors. In particular, its sugar pucker is influenced by the interaction with the 3′-base, in this case C

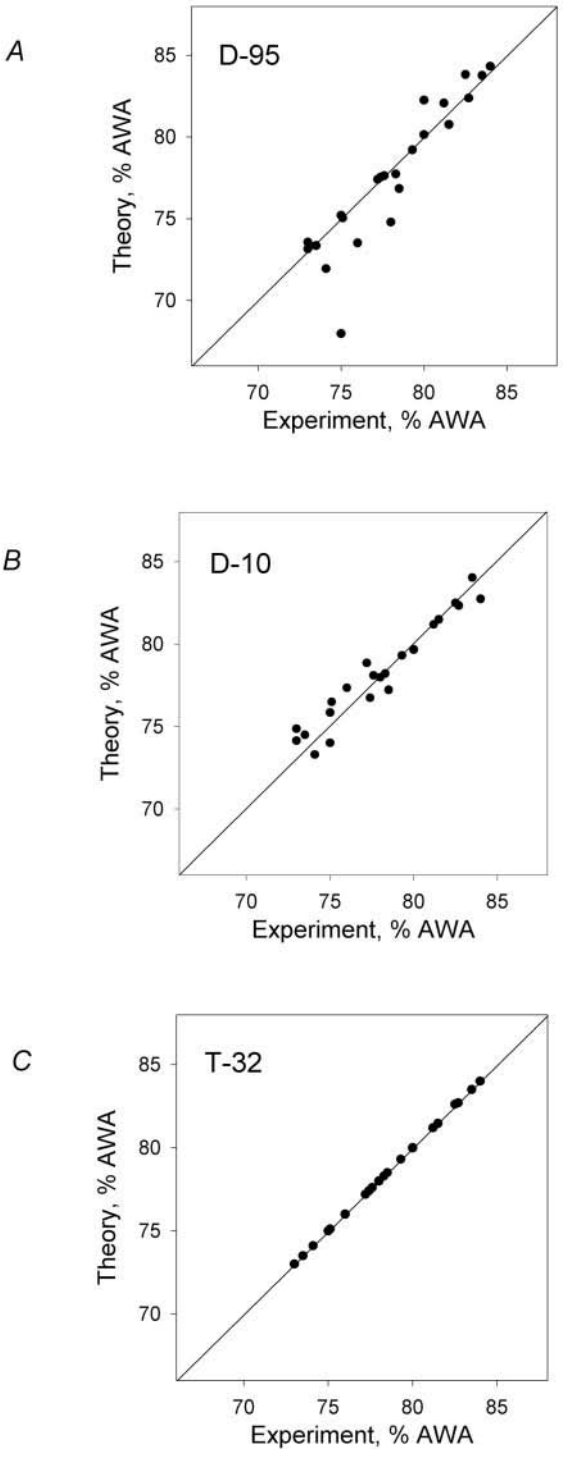

FIGURE 4 Comparison of the experimental values of the B↔A transition midpoints with the theoretical predictions for the three models: dimeric D-95 (*A*) and D-10 (*B*), and trimeric T-32 (*C*) (see Table 1). Correlation between theory and experiment increases from D-95 (correlation coefficient, $r^2 = 0.83$) to D-10 ($r^2 = 0.93$) to T-32 ($r^2 = 0.99$). Dimeric model D-10/12 (Table 1) gives deviations similar to those obtained for the model D-10.

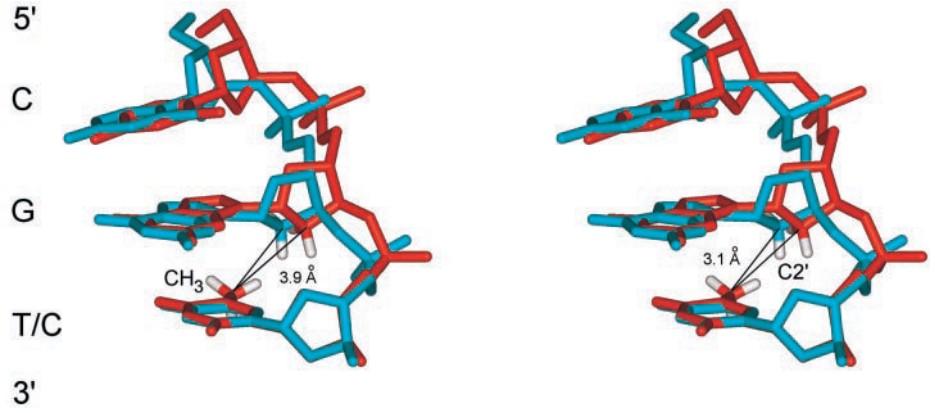

FIGURE 5 Stereoview of the hydrophobic interaction between the thymine methyl group and sugar ring of the preceding nucleotide (C2′H_2 group). Two trimeric steps are shown (minor groove view): 5′-CGC-3′ (*cyan*) (Woods et al., 2000; NDB entry bd0029, Berman et al., 1992) and 5′-CGT-3′ (*red*) (Rozenberg et al., 1998; NDB entry bd0001). The two trimers are superposed at the 3′-pyrimidine aromatic rings. Assuming the van-der-Waals radius of carbon to be 1.75 Å, one can see from the given distances that the GC step in the given configuration cannot accommodate the methyl group. To relieve the clash between CH_3 and C2′ groups, the guanine pseudorotation phase angle in the CGT trimer (P = 115.8°) is diminished compared with CGC (P = 161.5°). As a result, the C2′ group moves into the minor groove and the CH_3-C2′ distance increases from 3.1 to 3.9 Å.

or T. The C2′(H_2) group of guanine interacts either with C5H of cytosine or with the bulky methyl group of thymine. In the case of thymine, the tight contact between the methyl group and sugar constrains the DNA conformation (Ulyanov and Zhurkin, 1982; Hunter, 1993). One of the ways to relieve this possible steric conflict is to decrease the pseudorotation phase angle P by ~40°, as shown in Fig. 5. As a consequence, the C2′(H_2) group of guanine in CGT is exposed in the minor groove, thus increasing the hydrophobic surface of the central deoxyribose. Obviously, this rearrangement is important for the B-A equilibrium. The trimeric model can automatically account for this factor, and the free energy values ΔG_{BA} are adjusted correspondingly for the CGC and CGT trimers. By contrast, interactions with both the 3′-end and 5′-end neighbors are ignored in the dimeric models. Apparently, this oversimplification is responsible for a relatively poor fit to the experimental data with the dimeric models (Table 1, Fig. 5).

Stereochemical origins of the sequence-dependent B/A-philicity

It has been known for a while that poly(dA)·poly(dT) is profoundly resistant to the B→A transition (Arnott and Selsing, 1974; Ivanov and Krylov, 1992). From our analysis, it also follows that the steps AA:TT and AAN:N′TT are exceptionally B-philic (Tables 2 and 3). Why are these motifs different from the other ones?

In our opinion, the B-philicity of AAN:N′TT steps is related to the interactions involving the thymine methyl group, CH_3(T) (Fig. 6). In B-DNA, the CH_3(T) group is in direct contact with the C2′(H_2) group of the preceding deoxyribose, which may serve as a stabilizing factor in a hydrophilic environment (Ulyanov and Zhurkin, 1982). Moreover, when two consecutive thymines are in the same strand, their methyl groups, together with the mentioned methylene sugar groups and C6(T) atoms, form a compact hydrophobic cluster, thereby further stabilizing the B conformation (Fig. 6 *B*). We suggest this effect may be one of the reasons for the B-philicity of the TT-containing steps, along with other factors, such as the hydration spine in the minor groove (Drew and Dickerson, 1981). These two factors may be interrelated (e.g., thymine-sugar interactions in the major groove may lead to the negative roll, further stabilizing the hydration spine in the minor groove and vice versa).

To develop this model in further detail at the atomic level, one must know the structure of A-DNA for the so-called adenine tracts (the stretches of adenines in one strand, and thymines in the other strand). Unfortunately, the "pure" A-DNA crystals do not contain even a single AA:TT step, thus hindering a direct comparison between the B and A forms. The only available relevant structures either contain the modified backbone, N3′ → P5′ phosphoramidate instead of the phosphate groups (Tereshko et al., 1998) or are stabilized in the presence of actinomycin D (Kamitori and Takusagawa, 1994). One of the latter configurations is shown in Fig. 6 *A*, where the CH_3(T) groups are apparently closer to each other than in the B form (Fig. 6 *B*). The distance between the two methyls becomes as low as 3.7 A, which probably serves as a destabilizing factor for the A form. Besides, in A-DNA, the C2′(H_2) sugar groups are not in the immediate vicinity of the thymine methyl groups, as they are directed toward the minor groove (Fig. 6 *A*). This disruption of the hydrophobic cluster upon the B→A transition may be an additional factor stabilizing the B form in the hydrophilic environment.

In addition to the AA:TT step, the above CH_3-CH_2 interactions were suggested to favor the B conformation in the

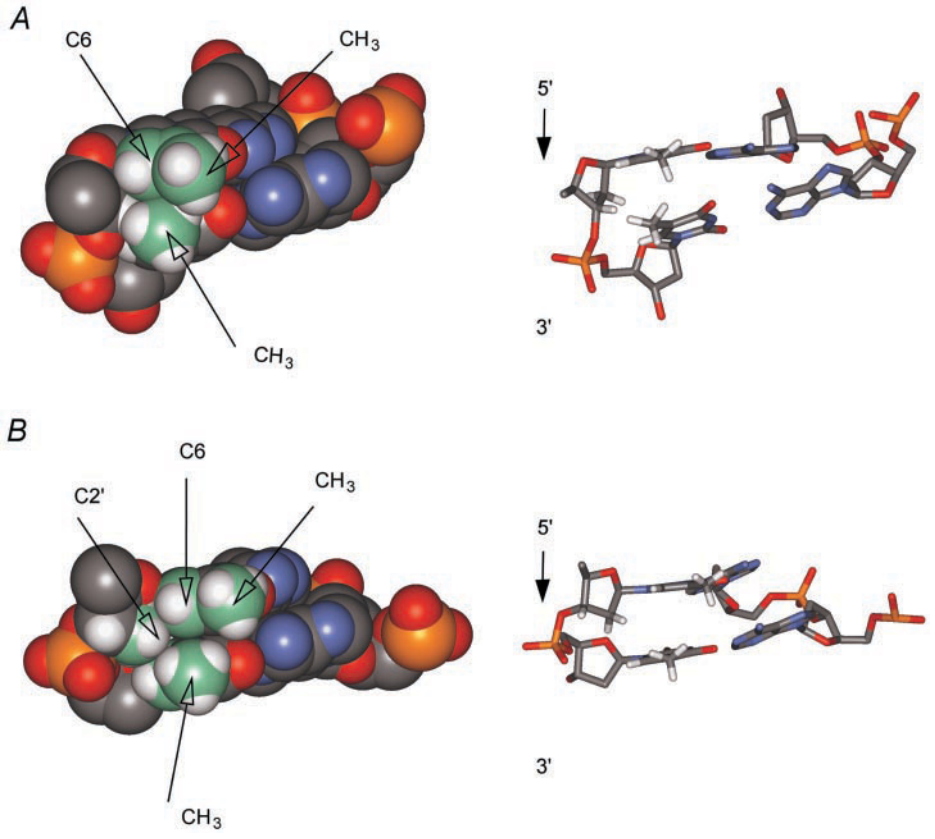

FIGURE 6 Interactions of the thymine methyl groups in the major groove of A form (*A*) and B form (*B*). The CPK representations are on the left (with the interacting carbons shown in *green*), and the skeleton ones on the right. The two dimers AA:TT are taken respectively from Kamitori and Takusagawa (1994; NDB entry, adh054) and Drew and Dickerson (1981; NDB entry, bdl001). Notice that in the A form, the sugar C2′H_2 group faces the minor groove, whereas in the B form, it interacts directly with the CH_3 group from the adjacent thymine.

case of AT:AT, the only dimer containing two such contacts (Ulyanov and Zhurkin, 1982). This prediction is consistent with the results obtained here (Table 2). In addition, the substitution of 5-methylcytosine for cytosine increases the alcohol concentration necessary for the B→A transition (Shchelkina et al., 1988), in agreement with the interpretation just given. To further assess the role of the methyl groups in stabilizing the B form, direct measurements are necessary with deoxyuridine substituted for thymine (work in progress).

Formation of the cross-strand hydrogen bonds between the bases from neighboring pairs may be another factor stabilizing the B form in a sequence-specific way (Nelson et al., 1986). For example, comparison of the MN and NM dimers shows that the one having the stronger potential for such H-bonding is more B-philic. There are two pairs of dimers to be considered: CA:TG versus AC:GT and AG:CT versus GA:TC (AT, TA, GC, and CG cannot form cross-strand H-bonds; AA:TT and GG:CC are excluded because for them MN = NM). The CA:TG dimer potentially can form two bifurcated H-bonds (in both grooves), whereas the AC:GT can form only one H-bond in the major groove (Mohan and Yathindra, 1991). Accordingly, the CA is more B-philic than AC (Table 2). The AG:CT can form one H-bond in the minor groove, and the GA:TC can form none. Again, the AG dimer turns out to be more B-philic than GA (Table 2).

Comparison with other works

Next we compare our results with other recent computational studies. Mazur et al. (1989) have shown that the base-base interactions in the GG:CC, GC, AT, and TA dimeric steps favor the A-like conformation, whereas the B form is more favorable for the AA:TT step. This agrees in general with our data except for the B-philic AT step, where the sugar-base interactions, not considered by Mazur et al. (1989), are likely to be important for stabilization of the B form (see above).

Other computations also demonstrate the importance of the atomic details for modeling the interactions between backbone and bases. For example, Hunter (1993) represented the deoxyriboses by methyl groups at the C1′ positions and considered various "conformational strategies" to avoid the steric clashes between the thymine methyl group

and the 5′-neighboring "sugar" modeled in such a way. He suggested that these clashes "block A-DNA conformations in AX/XT steps." Although this assessment is consistent with the B-philicity of the AA:TT and AT steps, it contradicts the pronounced A-philicity of the AC:GT observed here (Table 2). Using another simplified "mechanistic" model, Suzuki et al. (1996) arrived at the opposite conclusion, namely, that DNA has to switch from B-DNA to "a very A-like conformation" to avoid the collision between the thymine methyl groups and the backbone in the AT step. Note, however, that all-atom calculations (Ulyanov and Zhurkin, 1982) demonstrated the feasibility of the B-like conformation for the AT step. This earlier result is entirely consistent with our present data, indicating that AT is one of the most B-philic steps (Table 2).

On the other hand, a simple analysis of the base-base interactions turns out to be sufficient for the limited case of GC pairs. In addition to Mazur et al. (1989) mentioned above, Hunter (1993) has also correctly evaluated the B/A-philicity scale for the GC-containing dimers. He has shown that the base-base electrostatic interactions are repulsive for the GC, CG, and GG:CC steps in the "canonical" B-conformation with the helical axis going through the geometric centers of the base pairs. In the case of GC and GG:CC steps, these interactions become attractive for the B→A transition, because the negative slide (characteristic of the A form) brings the negatively charged N7 and O6 atoms of guanine into the vicinity of the amino group of cytosine. This strategy does not work for the CG step, however, where the more favorable slide values are predicted to be positive (Hunter, 1993). These simple rules agree with the A-philicity of the GG:CC and GC steps, and the B-philicity of the CG step, revealed in the present study.

In principle, the B/A-philicity can be evaluated from the accessible surface areas and solvation free energies calculated for numerous DNA fragments in the B and A conformations. This approach was used by Basham et al. (1995) who derived the trimeric code for the A propensity based on calculations of the accessible surface areas for the 32 trimers in "canonical" A and B conformations. This code does not fit the experimental data on the B↔A transitions used here and does not correlate with the trimeric scale deduced from these data (Fig. 7). There are two possible reasons for such an inconsistency. First, using "canonical" A and B conformations for all trimers, the authors ignored the sequence-specificity discussed above. Second, Basham et al. (1995) used the set of atomic solvation parameters corresponding to pure water conditions, whereas the B↔A transition takes place in a rather dehydrated and nonpolar environment (80% of RH corresponds to ~70% of ethanol).

Molecular dynamics (MD) simulations of various duplexes with explicit solvent would be helpful in elucidating the mechanism(s) of the sequence dependent B/A-philicity of DNA at the atomic level. Note that the MD studies of short duplexes in water-ethanol mixtures reproduce the

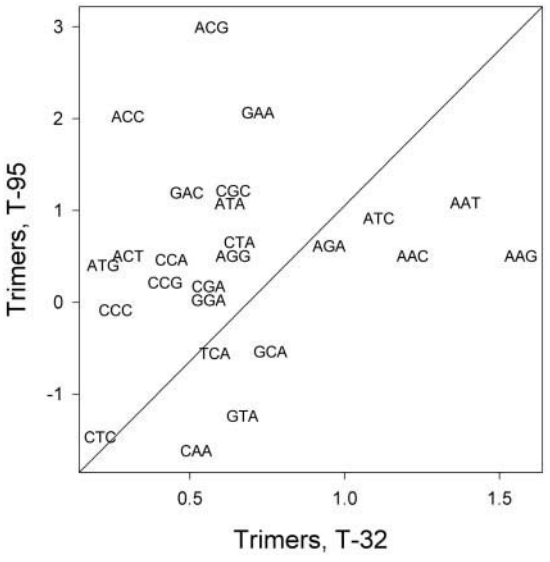

FIGURE 7 Absence of correlation between the ΔG_{BA} values (in kcal/mol) for the two trimeric models: T-32 obtained here (Table 3) and T-95, the model proposed by Basham et al. (1995). Several trimers are omitted for clarity. The centers of three-letter labels are placed at the corresponding data points on the graph.

B→A transition in general agreement with experiment (Cheatham et al., 1997; Feig and Pettitt, 1998; Sprous et al., 1998; Jayaram et al., 1998). The calculated B-A equilibrium depends, however, on the selection of the force field, e.g., AMBER or CHARMM (Feig and Pettitt, 1998; Sprous et al., 1998). Hopefully, further development of the potential functions will resolve this inconsistency (for review, see Olson and Zhurkin, 2000).

The influence of solvation on DNA conformation was studied thoroughly by Feig and Pettitt (1999). The distinct hydration patterns were found to contribute to the stabilization of the A and B forms, the A form being less hydrated than the B form. At the same time, the exposed sugar rings in A-DNA had more hydrophobic contacts with waters than the sugars in B-DNA, in particular around the A:T base pairs. This effect may be responsible for the resistance of the A/T rich sequences to undergoing the B→A transition. Furthermore, this finding supports the idea that alcohol shifts the B-A conformational equilibrium toward A-DNA, at least partially, by replacing unfavorable water molecules in the vicinity of the DNA hydrophobic surface. In addition, the detailed MD simulations of several DNA oligomers, made recently by McConnell and Beveridge (2000), revealed a remarkable complementarity between the DNA conformation and its solvation shell. It was demonstrated that the delicate balance of electrostatic components of internal DNA free energy, hydration, and electrostatic effects of counterion atmosphere is of major importance for the A- and B-DNA conformational stability.

Finally, Cheatham et al. (1998) computed the free-energy differences between the B-like and A-like conformations for

the G_{10}:C_{10} and A_{10}:T_{10} duplexes, which proved to be close to the experimental values presented here. It would be particularly interesting to deduce the hydration energies of various DNA moieties (e.g., bases, sugar rings, phosphates, and the CH_3—CH_2 clusters discussed above) directly from MD trajectories. Such an approach, based on a comparison of MD simulations with the experimental data for DNA oligomers of different sequences, could be used to explore sequence-dependent B/A-philicity in alcohol/water mixtures at the atomic level.

Implications of the model: genomic analysis

The three most studied DNA deformations (bending, twisting, and B↔ A transition) are involved in protein-DNA recognition, hence it is important to elucidate the interrelations between them. The DNA bending and twisting are strongly correlated (for review, see Olson and Zhurkin, 2000). The B↔A transition is also related to the other two deformations. For example, the local B→A transition is accompanied by major groove bending (Calladine and Drew, 1984). So, one would expect that the sequence-specific B/A code derived here (Table 2) could be linked to the "bendability code" (Mengeritsky and Trifonov, 1983; Dickerson, 1998; Olson et al., 1998). In other words, the A-philic dimers would be major-philic (i.e., easily bending into the major groove), whereas the B-philic dimers would be minor-philic.

For several (but not all) dimers this seems to be the case: the AA:TT and AT dimers are B-philic and minor-philic; the GG:CC and TA are A-philic and major-philic. By contrast, the dimers CA:TG, AG:CT, and CG are B-philic but at the same time prefer bending into the major groove (Olson et al., 1998). Thus, the dimeric B/A code derived here (Table 2) differs from the "bendability code." Similarly, the B/A code cannot be reduced to the DNA "twisting code" (Kabsch et al., 1982; Gorin et al., 1995).

There is one feature, however, that these two codes have in common. For the dimers MN and NM (AT-TA, AC-CA, AG-GA, CG-GC) the ΔG_{BA} energies alternate between high and low values, just as the DNA twist in solution (Kabsch et al., 1982). This trend prevents formation of long A-philic or B-philic DNA stretches (except A_n:T_n or G_n:C_n) in the same way as the DNA "twisting code" prevents accumulation of over- or undertwisting along DNA (Kabsch et al., 1982). On the other hand, the short 3- to 4-bp-long A-philic stretches could serve as signals for protein binding (Lu et al., 2000; Tolstorukov et al., 2001). Therefore, it would be interesting to apply the B/A codes established above for the analysis of genomic sequences in the context of the B↔A transition. To this aim, we suggest using the trimeric code, T-32 (Table 3), because it estimates the free energy change, ΔG_{BA}, for a given oligonucleotide more precisely than do dimeric codes.

So far, the sequence specificity of the B↔A transition remains less well understood than those for bending or twisting. Whereas the latter two can be effectively (but not completely) reduced to the local small-scale deformations within the B family, the B↔A transition involves the sugar switch leading to significant changes in mutual orientations of the interacting groups (bases, sugars, and phosphate moieties). Due to the complexity of these interactions and the flexibilities of the moieties involved, comparisons between the computer calculations and the experimental data are quite complicated, and we are only beginning to decipher the rules governing the B↔A transition at the atomic level.

One of the simple rules established here is stabilization of the B form by hydrophobic interactions between the thymine methyl and the sugar methylene groups (in the dimeric steps AA:TT, AT, and AG:CT). In the framework of the "RNA world" concept (Gesteland et al., 1998), one can hypothesize that the thymine methyl group was incorporated into DNA to stabilize the B form, when the DNA began playing a role of the long-term repository of genetic information. (DNA is chemically more stable than RNA, and B-form DNA has a shorter persistence length than A-form RNA, which is why B-DNA is more suitable for the storage of information in a limited space inside the cell.)

The empirical parameter AWA used in this study accounts implicitly not only for DNA hydration, but also for the interactions with alcohol, ions, and other solvent components. Our approach accounts for all of these interactions in a phenomenological way, treating the DNA-solvent system as a "black box" and the AWA parameter as the "thermodynamic driving force" of the B↔A transition. Hopefully, this simple experiment-based model will be applicable to the DNA-protein complexes as well. Certainly, more data is necessary to verify this. So far, the trimeric B/A-philicity scale established here proves to be remarkably consistent with the frequency of occurrence of A forms in the DNA-protein cocrystals. Therefore, we anticipate that this sequence-dependent B/A scale would be useful for assessing the ease of the B→A transition in large nucleoprotein assemblages, such as the transcription and replication complexes.

The paper is dedicated to the memory of Dr. Ludmila E. Minchenkova. The authors are grateful to Drs. D.Y. Krylov, W.K. Olson, A.K. Shchyolkina and N.B. Ulyanov for valuable discussions and constructive criticism, as well as to W.K. Olson and A. Colasanti for providing their database on the DNA structural parameters in crystals.

REFERENCES

Arnott, S., and E. Selsing. 1974. Structures for the polynucleotide complexes poly(dA) with poly(dT) and poly(dT) with poly(dA) with poly(dT). *J. Mol. Biol.* 88:509–521.

Basham, B., G. P. Schroth, and P. S. Ho. 1995. An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. U.S.A.* 92:6464–6468.

Becker, M., and Z. Wang. 1988. B-A transitions within a 5 S ribosomal RNA gene are highly sequence-specific. *J. Biol. Chem.* 264: 4163–4167.

Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63: 751–759.

Calladine, C. R., and H. R. Drew. 1984. A base-centered explanation of the B-to-A transition in DNA. *J. Mol. Biol.* 178:773–782.

Cheatham, T. E. III, M. F. Crowley, T. Fox, and P. A. Kollman. 1997. A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. *Proc. Natl. Acad. Sci. U.S.A.* 94:9626–9630.

Cheatham, T. E. III, J. Srinivasan, D. A. Case, and P. A. Kollman. 1998. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dynam.* 16:265–280.

Cheetham, G. M., and T. A. Steitz. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science.* 286:2305–2309.

Cooney, A., and K. W. Morcom. 1988. Thermodynamic behaviour of mixtures containing fluoroalcohols: I. (Water + 2,2,2-trifluoroethanol). *J. Chem. Therm.* 20:735–741.

Dickerson, R. E. 1998. DNA bending:the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* 26:1906–1926.

Ding, J., K. Das, Y. Hsiou, S. G. Sarafianos, A. D. J. Clark, A. Jacobo-Molina, C. Tantillo, S. H. Hughes, and E. Arnold. 1998. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* 284:1095–1111.

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

El Hassan, M. A., and C. R. Calladine. 1997. Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Phil. Trans. R. Soc. Lond. A.* 355:43–100.

Feig, M., and B. M. Pettitt. 1998. Structural equilibrium of DNA represented with different force fields. *Biophys. J.* 75:134–149.

Feig, M., and B. M. Pettitt. 1999. Modeling high-resolution hydration patterns in correlation with DNA sequence and conformation. *J. Mol. Biol.* 286:1075–1095.

Foloppe, N., and A. D. MacKerell. 1999. Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys. J.* 76: 3206–3218.

Franklin, R. E., and R. G. Gosling. 1953. The structure of sodium thymonucleate fibers: I. The influence of water content. *Acta Crystallogr.* 6:673–677.

Gesteland, R. F., T. R. Cech, and J. F. Atkins. 1998. The RNA World, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Gorin, A. A., N. B. Ulyanov, and V. B. Zhurkin. 1990. S-N transition of the sugar ring in B-form DNA. *Mol. Biol.* 24:1036–1047.

Gorin, A. A., V. B. Zhurkin, and W. K. Olson. 1995. B-DNA twisting correlates with base pair morphology. *J. Mol. Biol.* 247:34–48.

Hunter, C. A. 1993. Sequence-dependent DNA structure: the role of base stacking interactions. *J. Mol. Biol.* 230:1025–1054.

Ivanov, V. I., and D. Y. Krylov. 1992. A-DNA in solution as studied by diverse approaches. *Methods Enzymol.* 211:111–127.

Ivanov, V. I., D. Y. Krylov, and E. E. Minyat. 1985. Three-state diagram for DNA. *J. Biomol. Struct. Dynam.* 3:43–55.

Ivanov, V. I., and L. E. Minchenkova. 1995. The A-form of DNA: in search of biological role (a review). *Mol. Biol.* 28:780–788.

Ivanov, V. I., L. E. Minchenkova, G. Burckhardt, E. Birch-Hirschfeld, H. Fritzsche, and C. Zimmer. 1996. The detection of B-form/A-form junction in a deoxyribonucleotide duplex. *Biophys. J.* 71:3344–3349.

Ivanov, V. I., L. E. Minchenkova, B. K. Chernov, P. McPhie, S. Ryu, S. Garges, A. M. Barber, V. B. Zhurkin, and S. Adhya. 1995. CRP-DNA complexes: inducing the A-like form in the binding sites with an extended central spacer. *J. Mol. Biol.* 245:228–240.

Ivanov, V. I., L. E. Minchenkova, E. E. Minyat, M. D. Frank-Kamenetskii, and A. K. Schyolkina. 1974. The B to A transition of DNA in solution. *J. Mol. Biol.* 87:817–833.

Ivanov, V. I., L. E. Minchenkova, E. E. Minyat, and A. K. Schyolkina. 1983. Cooperative transitions in DNA with no separation of strands. *Cold Spring Harbor Symp. Quant. Biol.* 47:243–250.

Ivanov, V. I., V. B. Zhurkin, S. K. Zavriev, Yu. P. Lysov, L. E. Minchenkova, E. E. Minyat, M. D. Frank-Kamenetskii, and A. K. Schyolkina. 1979. Conformational possibilities of double-helical nucleic acids: theory and experiment. *Int. J. Quantum Chem.* 16:189–201.

Jayaram, B., D. Sprous, M. A. Young, and D. L. Beveridge. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* 120:10629–10633.

Kabsch, W., C. Sander, and E. N. Trifonov. 1982. The ten helical twist angles of B-DNA. *Nucleic Acids Res.* 10:1097–1104.

Kamath, S., M. H. Sarma, V. B. Zhurkin, C. J. Turner, and R. H. Sarma. 2000. DNA bending and sugar switching. *J. Biomol. Struct. Dynam.* 11:317–325.

Kamitori, S., and F. Takusagawa. 1994. Multiple binding modes of anticancer drug actinomycin D: x-ray, molecular modeling, and spectroscopic studies of $d(GAAGCTTC)_2$-actinomycin D complexes and its host DNA. *J. Am. Chem. Soc.* 116:4154–4165.

Kiefer, J. R., C. Mao, J. C. Braman, and L. S. Beese. 1998. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature.* 391:304–307.

Lindsay, S. M., S. A. Lee, J. W. Powell, T. Weidlich, C. DeMarco, G. D. Lewen, N. J. Tao, and A. Rupprecht. 1988. The origin of the A to B transition in DNA fibers and films. *Biopolymers.* 27:1015–1043.

Lu, X-. J., Z. Shakked, and W. K. Olson. 2000. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* 300:819–840.

Maleev, V. Y., M. A. Semenov, A. I. Gasan, and V. A. Kashpur. 1993. Physical properties of the DNA-water system. *Biophysics.* 38:789–811.

Malenkov, G. G., and K. A. Minasyan. 1977. Transition of DNA from the B to A form in aqueous solutions of nonelectrolytes. *Mol. Biol.* 11: 268–274.

Malenkov, G., L. Minchenkova, E. Minyat, A. Schyolkina, and V. Ivanov. 1975. The nature of the B-A transition of DNA in solution. *FEBS Lett.* 51:38–42.

Mazur, J., A. Sarai, and R. L. Jernigan. 1989. Sequence dependence of the B-A conformational transitions of DNA. *Biopolymers.* 28:1223–1233.

McConnell, K. J., and D. L. Beveridge. 2000. DNA structure: what's in charge? *J. Mol. Biol.* 304:803–820.

Mengeritsky, G., and E. N. Trifonov. 1983. Nucleotide sequence-directed mapping of the nucleosomes. *Nucleic Acids Res.* 11:3833–3851.

Minchenkova, L. E., A. K. Schyolkina, B. K. Chernov, and V. I. Ivanov. 1986. CC/GG contacts facilitate the B to A transition of DNA in solution. *J. Biomol. Struct. Dynam.* 4:463–476.

Mohan, S., and N. Yathindra. 1991. Studies on the cross strand hydrogen bonds in DNA double helices. *J. Biomol. Struct. Dynam.* 9:113–126.

Nekludova, L., and C. O. Pabo. 1994. Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes. *Proc. Natl. Acad. Sci. U.S.A.* 91:6948–6952.

Nelson, H. C. M., J. T. Finch, B. F. Luisi, and A. Klug. 1986. The structure of an oligo(dA):oligo(dT) tract and its biological implications. *Nature.* 330:221–226.

Nishimura, Y., C. Torigoe, and M. Tsuboi. 1986. Salt induced B-A transition of poly(dG)×poly(dC) and the stabilization of A form by its methylation. *Nucleic Acids Res.* 14:2737–2748.

Olson, W. K., A. A. Gorin, X.-J. Lu, L. M. Hock, and V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* 95:11163–11168.

Olson, W. K., and V. B. Zhurkin. 2000. Modeling DNA deformations. *Curr. Opin. Struct. Biol.* 10:286–297.

Rozenberg, H., D. Rabinovich, F. Frolow, R. S. Hegde, and Z. Shakked. 1998. Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl. Acad. Sci. U.S.A.* 95: 15194–15199.

Shakked, Z., G. Guzikevich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak, and P. B. Sigler. 1994. Determinants of repressor/operator recognition from the structure of the *trp* operator binding site. *Nature.* 368:469–473.

Shchelkina, A. K., L. E. Minchenkova, V. I. Ivanov, V. V. Butkus, and A. A. Yanulaitis. 1988. B-A and B-Z transitions in deoxyoligoduplexes containing 4- and 5-methylcytosine. *Mol. Biol.* 22:1243–1250.

Sprous, D., M. A. Young, and D. L. Beveridge. 1998. Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: theoretical considerations of the A-B transition. *J. Phys. Chem. B.* 102:4658–4667.

Suzuki, M., N. Yagi, and J. T. Finch. 1996. Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett.* 379:148–152.

Tereshko, V., S. Gryaznov, and M. Egli. 1998. Consequences of replacing the DNA 3oxygen by an amino group: high-resolution crystal structure of a fully modified N3′→P5′ phosphoramidate DNA dodecamer duplex. *J. Am. Chem. Soc.* 120:269–283.

Timsit, Y. 1999. DNA structure and polymerase fidelity. *J. Mol. Biol.* 293:835–853.

Tolstorukov, M. Ye., R. L. Jernigan, and V. B. Zhurkin. 2001. Protein-DNA minor groove recognition. *J. Biomol. Struct. Dynam.* 18:944–945.

Ulyanov, N. B., and V. B. Zhurkin. 1982. Flexibility of complementary dinucleoside phosphates: a Monte Carlo study. *Mol. Biol.* 16:857–867.

Woods, K. K., L. McFail-Isom, C. C. Sines, S. B. Howerton, R. K. Stephens, and L. D. Williams. 2000. Monovalent cations sequester within the A-tract minor groove of $[d(CGCGAATTCGCG)]_2$. *J. Am. Chem. Soc.* 122:1546–1547.