DOI: 10.1002/chem.200601120

On the Nature of DNA-Duplex Stability

Jan Řezáč and Pavel Hobza*[a]

Abstract: The unwinding free energy of 128 DNA octamers was correlated with the sum of interaction energies among DNA bases and their solvation energies. The former energies were determined by using the recently developed density functional theory procedure augmented by London dispersion energy (RI-DFT-D) that provides accurate hydrogen-bonding and stacking energies highly comparable with CCSD(T)/complete basis set limit benchmark data. Efficient tight-binding DFT covering dispersion energy was also used and yielded satisfactory results. The latter method can be used for extended systems. The solvation energy was determined by using a C-PCM continuum solvent at HF level calculations. Various models were adopted to correlate theoretical energies with experimental unwinding free energies. Unless all energy components (hydrogen-bonding, intra- and interstrand-stacking, and solvation energies) were included and weighted individually, no satisfactory correlation resulted. The most advanced model yielded very close correlation (RMSE = 0.32 kcal mol^{-1}) fully comparable with the entirely empirical correlation introduced in the original paper.[3] Analysis of the theoretical results shows the importance of inter- and intramolecular stacking energies, and especially the latter term plays a key role in determining DNA-duplex stabilization.

Keywords: ab initio calculations · density functional calculations · DNA stability · hydrogen bonds · solvation energy

Introduction

The DNA double-helical structure is important for the storage and transfer of genetic information. The structure results from different noncovalent energy contributions of various building blocks of DNA, among which the interactions of nucleic acid (NA) bases play a dominant role. The NA bases are polar, aromatic heterocycles that interact either through planar hydrogen bonds or vertical π–π interactions, resulting in two structural motifs; planar hydrogen bonding and π stacking. Both motifs are important not only in determining the architecture of nucleic acid, but also in a much more general sense. The basic question is: what is the relative strength of these interactions? It was believed for a long time that specific hydrogen-bonding interactions originating mainly from electrostatic effects were dominant, whereas nonspecific stacking originating from London dispersion effects was considered to be energetically much less significant. Recent accurate, correlated ab initio quantum chemical calculations have revealed[1] that stacking can be associated with surprisingly large stabilization energies, comparable with those of strong hydrogen bonding. It was even shown that in the case of adenine-rich DNA sequences, the hydrogen-bonding and stacking interaction energies are comparable.[2] The stabilization energies of NA base pairs are certainly important for DNA stability, as unwinding is associated with the complete loss of hydrogen bonding and interstrand stacking, and with a partial loss of intrastrand stacking. The loss of hydrogen bonding has one additional, very important consequence. The polar NA bases are now exposed to water, and because dipole moments of NA are different ($\mu(\text{guanine}) \approx \mu(\text{cytosine}) > \mu(\text{thymine}) > \mu(\text{adenine})$), we obtain different guanine···cytosine and adenine···thymine solvation energies. Evidently, the DNA-duplex stability is proportional not only to the stabilization energies of NA bases, but also to their solvation/desolvation energies.

Recently, Doktycz et al.[3] determined melting temperatures and unwinding free energies for 140 octamer duplexes,

[a] J. Řezáč, P. Hobza
Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic and
Center for Biomolecules and Complex Molecular Systems
Flemingovo nám. 2, 16610 Prague 6 (Czech Republic)
Fax: (+420) 220-410-320
E-mail: pavel.hobza@uochb.cas.cz

Supporting information for this article is available on the WWW under http://www.chemeurj.org/ or from the author: Table of the sums of interaction energies (E_h, E_i, and E_s), calculated by using both RI-DFT-D and DFTB-D, for the set of 128 octamers, along with DNA sequence information.

and predicted their stabilities on the basis of various purely empirical correlation models, including the nearest-neighbor model introduced by Breslauer et al.[4] The correlation parameters (from 12 to 42, depending on the model used), not reflecting any physical nature of binding, were determined from a set of 128 duplexes and were later applied to a validation set of 12 duplexes. The observed unwinding free energies of these 12 duplexes were predicted with an RMSE (root mean square error) of 0.35 kcal mol^{-1}.

The aim of the present study was to analyze various contributions to the overall unwinding free energy of duplex formation on the basis of accurate quantum chemical calculations of hydrogen-bonded, intrastrand- and interstrand-stacked NA base pairs, and solvation energies of NA bases. We are certainly aware that we cannot determine the real unwinding free energy of the DNA duplex. Our model is limited to only NA bases, with the remaining components of the DNA duplex (sugars, phosphates) being ignored. We also combine the interaction energies of various NA base pairs with solvation free energies. The sum of both energies will be first correlated with experimental values of unwinding free energies. The correlation coefficients calculated will be used later for subsequent prediction. The main advantage of the present approach is that all interactions of NA bases (including their solvation) are correctly and accurately described. The reason for introducing scaling is that the real system is very complicated and it is beyond present possibilities to calculate unwinding free energies accurately. It is assumed that neglected factors (e.g., the role of sugars, phosphates, entropy) are either similar in all DNA duplexes (like the role of a backbone) and are, therefore, accounted for in a fitted constant, or are similar in a DNA duplex and single strands and are, thus, canceled. A third possibility is that their sum is proportional to the quantities considered (interaction energies of NA bases and solvation free energies of NA bases) and is, thus, included in the single scaling factor adopted.

The important advantage of the present procedure is that it can also be applied to more complicated, unusual, or unnatural DNA structures for which the empirical correlation suggested in references [3,4] cannot be used.

Strategy

Because the structures of octamer duplexes are unknown, we constructed them by using modeling software and optimized them by using molecular mechanics. The geometries of hydrogen-bonded, intra- and interstrand-stacked NA base pairs were taken from the double-helix geometries. Stabilization energies of all these pairs were calculated by using the density functional theory procedure augmented by London dispersion energy (RI-DFT-D).[5] For both hydrogen-bonded and stacked NA base pairs, this procedure yields very accurate stabilization energies comparable to benchmark CCSD(T)/complete basis set limit values. Additionally, stabilization energies were also determined by using the very fast semiempirical self-consistent charge–density-functional–tight-binding methods also augmented by London dispersion energy (SCC-DFTB-D).[6] This method, which is computationally much more feasible, also provides reliable stabilization energies for various structures of NA base pairs.

Although the evaluation of the stabilization energies requires only the knowledge of base-pair geometry (taken from duplex geometry), solvation/desolvation free energies are based on a knowledge of geometries of both duplexes and single strands. Because no experimental evidence concerning single-strand geometry exists, we decided to adopt the geometry from duplexes. Solvation free energies were determined by using the C-PCM procedure, which is based on considering a continuous solvent represented by a dielectric constant. This procedure is known to yield accurate solvation free energies of NA bases and base pairs.[7]

The final (effective) unwinding free energy was constructed as a weighted sum of eight hydrogen-bonding energies (E_h), 14 interstrand- (E_i) and 14 intrastrand- (E_s) stacked energies, and a contribution to solvation free energy. The latter term was determined as the difference between solvation free energies of AT and GC pairs multiplied by their occurrence in the particular sequence. Weighting coefficients were optimized to fit the calculated ΔG to the experimental value over the training set of 128 octamers. The relationship found was then tested on a validation set of 12 octamers.

Methods

Preparation of model duplexes: The structure of the DNA duplex was created by using the nucgen program, a part of the AMBER package.[8] Molecular mechanic (Cornell et al. potential[9]) optimization in implicit solvent was performed on the constructed structure to refine the sequence-dependent geometric properties. A generalized Born model (GBM),[10,11] considering an implicit solvent, implemented in the AMBER package, was used, with implicit treatment of counterions.[12] The concentration of the virtual salt was set to 0.1 mol^{-1}. The NA bases were then extracted from the optimized structure, and the glycosidic bond was terminated with hydrogen at an optimal distance.

Interaction energies: The pairwise interaction energies were calculated by using the RI-DFT-D procedure, which combines the DFT/TPSS/TZVP interaction energies and empirical London dispersion energy.[13] Interaction energies were determined as the difference between the energies of complex and of isolated subsystems; the basis set superposition error in the present treatment is small and can, therefore, be neglected. For computational treatment, including the parametrization of the dispersion-energy term, see the original paper.[13] The resolution of identity approximation within the DFT (RI-DFT) was used to improve efficiency,[14] as implemented in the Turbomole package.[15] Besides the RI-DFT-D, the semiempirical SCC-DFTB-D (self-consistent charge–density-functional–tight-binding) method with empirical-dispersion term[6] was also applied. The latter method is computationally very efficient, which makes it possible to determine energies for more-extensive systems as well.

Solvation: Because it is difficult to calculate absolute values of solvation free energy for both the whole duplex and the separate strands of DNA, we used an additive scheme based on the calculated $\Delta\Delta G^{solv}$ for one base pair. Firstly, the ΔG of solvation (ΔG^{solv}) was calculated for smaller model systems by using a continuum solvent model within quantum-mechanical calculations. The C-PCM method[16] implemented in a Gaussian package[17] based on the HF/6–31G(d) calculations was used with default optimized solvent parameters (UAHF radii).

The solvation free energy difference between the AT and GC pairs embedded in DNA was calculated as follows: model trimer duplexes (CGC/GCG and CTC/GAG) were prepared and neutralized by using the protocol described above. Solvation free energy both for the duplex and the separated strands was determined and the $\Delta\Delta G^{solv}$ was subsequently evaluated:

$$\Delta\Delta G^{solv} = \Delta G^{solv}(\text{duplex}) - \Delta G^{solv}(\text{strand A}) - \Delta G^{solv}(\text{strand B}) \quad (1)$$

The geometries of the single strands were considered to be the same as the geometry of the respective strand in the duplex, because no experimental evidence exists about their structure. We are aware that a single strand is more flexible than a duplex. However, the octamer strands are too short to allow any dramatic structural change. The assumption that the geometry remained unchanged was confirmed by pilot molecular dynamic simulations on a single strand; several ns simulations were performed, but practically no geometry changes were detected.

This approximation yielded the lower boundary of the solvation contribution, neglecting the increased exposition of bases due to the single-strand flexibility. By knowing the $\Delta\Delta G^{solv}$ for both trimers being considered, it was possible to calculate the difference between the CG and AT pairs ($\Delta\Delta\Delta G^{solv}$, denoted in the following text as DG for short) in the DNA double helix:

$$\mathrm{D}G^{solv}\ (\mathrm{CG}) = \Delta\Delta G^{solv}\ (\mathrm{CG}) - \Delta\Delta G^{solv}\ (\mathrm{AT}) \quad (2)$$

Because the solvation conditions at the end of the duplex are different, they must be considered separately. The scheme described above was also applied to these base pairs. As a model, dimers AC/GT and GC/GC were used, obtaining the value DG^{solv}(CG,end):

$$\mathrm{D}G^{solv}(\mathrm{CG,end}) = \Delta\Delta G^{solv}(\mathrm{CG,end}) - \Delta\Delta G^{solv}(\mathrm{CG}) \quad (3)$$

The total contribution of solvation to the ΔG, DG^{solv} (sequence) consists of the value dependent on CG contents, DG^{solv}(GC) and the correction for the end pairs, DG^{solv}(ends):

$$\mathrm{D}G^{solv}(\text{sequence}) = N_{CG} \times \mathrm{D}G^{solv}(\mathrm{CG}) \quad (4)$$

$$\mathrm{D}G^{solv}(\text{ends}) = N_{CG,end} \times \mathrm{D}G^{solv}(\mathrm{CG,end}) \quad (5)$$

in which N_{CG} is the number of CG pairs in the duplex and $N_{CG,end}$ is the number of the end CG pairs.

Estimating the ΔG of DNA-duplex unwinding: To estimate the ΔG of dissociation of the duplex in solution, we used calculated interaction energies and solvation energies (DG^{solv}). Because the DG^{solv} is a relative value describing the contribution of the basis only, a constant K of an unknown value was added. This constant also includes all the other sequence-independent contributions. The sums of interaction energies were scaled by the coefficients c to fit the range of experimental values. The final equation is:

$$\Delta G = K + c_h(E_h + \mathrm{D}G^{solv}(\text{sequence}) + \mathrm{D}G^{solv}(\text{ends})) + c_i(E_i + c_s)E_s \quad (6)$$

in which c_h is a coefficient describing the weight of hydrogen bonding corrected for the effect of solvation, whereas c_i and c_s describe the weight of interstrand and intrastrand stacking, respectively, in the duplex. Notice that the hydrogen-bonding term includes all the solvation-free-energy change related to duplex dissociation (assuming there is no change in the single-strand structure) and the stacking-related terms are not corrected.

The unknown values of K, c_h, c_i, and c_s were optimized to fit the experimentally measured values of unwinding free energies (the standard error of ΔG measurements[3] is reported to be 0.12 kcal mol^{-1}) for structures in the training set. These coefficients include all the contributions not covered by our calculations, such as the effect of entropy. Entropy disfavors complex formation, or in other words, it compensates the stabilization energy.[3,18] The RMSE between the calculated and experimental data was minimized by the fitting procedure. All the optimized values were constrained to be non-negative to conserve the physical meaning of the contributions.

The equation obtained was then applied to the validation set of 12 structures not included in the training set, and the $RMSE_{val}$ was determined.

Results and Discussion

Interaction energies: Table 1 shows averaged total hydrogen-bonding, interstrand- and intrastrand-stacking energies determined by using RI-DFT-D and DFTB-D methods for 128 octamer structures in a training set. Hydrogen-bonding energies determined by the latter method are underestimated by approximately 20%, whereas both stacking energies agree reasonably well with reference data. For the present purposes of approximating the ΔG, absolute values of interaction energies are not important, as the constant difference is accounted for by the fitted coefficients c and K. The correlation between the sum of pairwise interaction energies of a particular type of interaction obtained from both computational methods is very good ($R^2(E_h) = 0.9999$, $R^2(E_i) = 0.9903$, and $R^2(E_s) = 0{,}9949$). Evidently, the difference is negligible and the cheaper DFTB-D energies can be used safely in this application. From Table 1 it is clear that hydrogen-bonding energies are more important than stacking energies, but the order of magnitude is the same. In the case of the RI-DFT-D method, the stacking energies form 48% of the hydrogen-bonding energies, and for DFTB-D this ratio is even larger (67%, see Table 1). It should be mentioned that the former energies will be affected (reduced) by solvation much more than by stacked energies (see below). A detailed list of interaction energies for all 128 octamers is shown in Table 1 of the Supporting Information. Upon investigating these data, we found surprising complementarity in stacking energies—high values of intrastrand stacking are accompanied by low interstrand stacking and vice versa (see Figure 1). The sums of these energies are, thus, very similar. Notably, a similar conclusion based on a much lower number of DNA structures was reached in one of our previous papers.[19]

Table 1. Average total interaction energies (ΔE) in a set of 128 DNA octamers, and their values relative to hydrogen bonding (E^{rel}).

	ΔE [kcal mol^{-1}]		E^{rel} (% of H bonding)	
	RI-DFT-D	DFTB-D	RI-DFT-D	DFTB-D
H bonding	−204.2	−160.8	100	100
H bonding + solvation	−179.3	−135.9	88	85
interstrand stacking	−21.3	−20.5	10	13
intrastrand stacking	−78.0	−86.1	38	54

Solvation: The calculated values of the ΔG^{solv} for model trimer and dimer duplexes, as well as for the isolated base pair, are listed in Table 2. The difference between CG and AT pairs, DG^{solv}(CG), was calculated for each case. The GC

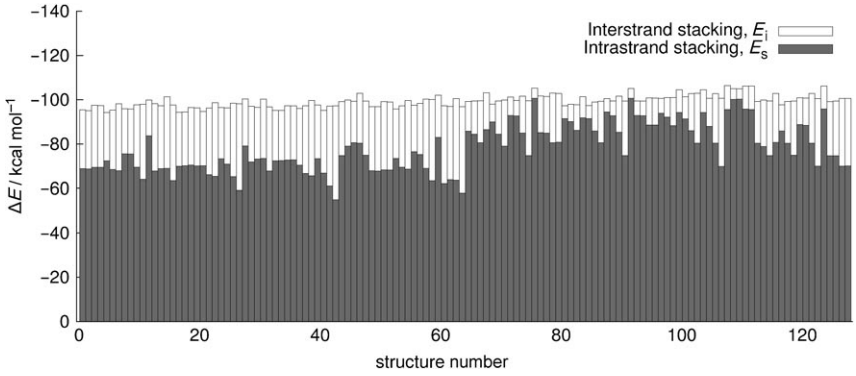

Figure 1. Complementarity of intrastrand and interstrand stacking energies in DNA octamer duplexes.

Table 2. Solvation-free-energy change [kcal mol^{-1}] upon dissociation calculated for different molecular fragments by using the C-PCM procedure.

Structure	$\Delta\Delta G^{solv}$	DG^{solv}(GC)
trimer CTC/GAG	58.7	
trimer CGC/GCG	63.8	5.1
dimer AC/GT	56.1	
dimer CC/GG	64.2	8.1
A/T	11.5	
C/G	25.6	14.1

pair in the gas phase is considerably more stable than the AT pair (by 17.2 and 14.9 kcal mol^{-1} for the RI-DFT-D and DFTB-D methods, respectively). Because the sum of dipole moments of guanine and cytosine is larger than that of adenine and thymine, the CG base pair is more destabilized by solvation than the AT pair. The difference in solvation free energy is largest for the isolated pairs that are fully exposed to water and amounts to about 14 kcal mol^{-1}. Thus, the mentioned substantial energy preference of the GC pair over the AT pair is almost completely compensated by solvation (the sum of the averaged interaction-energy difference and DG^{solv}(CG) equals 3.0 and 0.8 kcal mol^{-1} for the two methods, respectively). The value of the DG^{solv}(CG) term is reduced when the base pair is not free but is embedded inside the DNA. Bases are now less exposed to water, but the DG^{solv}(CG) value is still substantial (more than 8 and 5 kcal mol^{-1} for the dimer and trimer, respectively). The corresponding contribution of the CG pair located at the end of the helix (DG^{solv}(CG,end)) amounts to 3.0 kcal mol^{-1}. The average hydrogen-bonding energy in the set of octamers corrected for the solvation is listed in Table 1 and amounts to about 88 and 85% of the gas-phase interaction for the RI-DFT-D and DFTB-D methods, respectively.

Estimation of the change in unwinding free energy: To demonstrate the role of various energy components in the total change in unwinding free energy, we plotted the sums of interaction energies against the experimental ΔG (Figure 2). It is clear that all the components (Figure 2d) should be included to obtain the correlation with the unwinding free energy.

To improve the results and scale them to the range of experimental values, the fitting procedure (weighting each contribution individually) was subsequently used, as described above. We enhanced the model in five levels to demonstrate the importance of all the contributions (see Table 3). The quality of each model is judged by the correlation coefficient R,

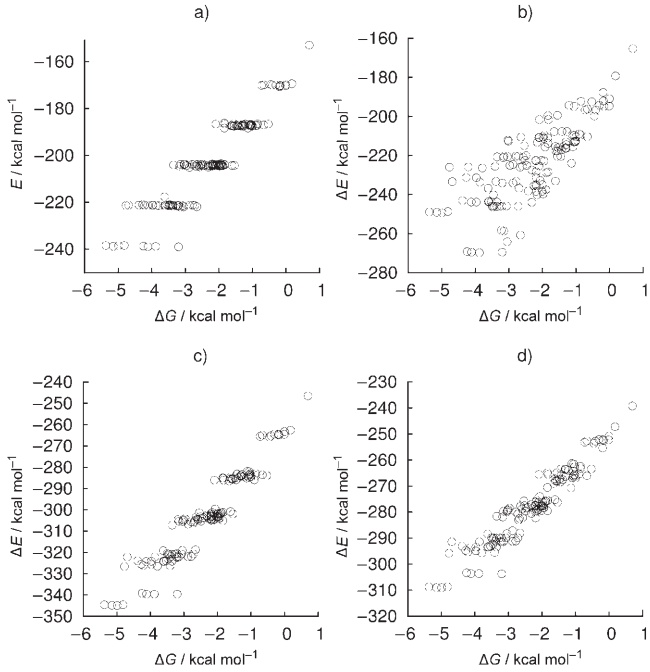

Figure 2. Sums of total interaction energies a) E_h, b) E_h+E_i, c) $E_h+E_i+E_s$ d) $E_h+E_i+E_s+DG^{solv}$ plotted against the ΔG of unwinding.

the RMSE for both training and validation sets, and the standard error for each optimized parameter (except level 2, at which a constrained optimization was applied).

Level 1: We considered only hydrogen bonding, corrected for solvation, and interstrand stacking. Intrastrand stacking and solvation were neglected. This level corresponded to the interaction of two strands, for which all intrastrand interactions were neglected, scaled by c_h and biased by K to fit the ΔG range. These values were optimized and the others were constrained: c_i was set to be equal to c_h, and c_s was equal to zero. No correlation was observed at this level ($R^2=0.36$), which indicates that although the contribution of interstrand interactions (hydrogen bonding plus interstrand stacking) is important, it is not sufficient to consider only this term.

Table 3. Fit of RI-DFT-D results to experimental unwinding energies [see Eq. (6)]. Optimized variables are shown in bold type. Standard errors of the coefficients are presented in parenthesis.

Level	1	2	3	4	5
K	**7.79** (±**1.21**)	**14.57**	**20.42** (±**0.66**)	**27.36** (±**1.50**)	**29.21** (±**1.4**)
c_h	**0.05** (±**0.01**)	**0.09**	**0.08** (±**0.003**)	**0.06** (±**0.004**)	**0.08** (±**0.005**)
c_i	$=c_h$	**0.00**	$=c_h$	**0.19** (±**0.02**)	**0.20** (±**0.02**)
c_s	0	0	**0.09** (±**0.003**)	**0.19** (±**0.02**)	**0.19** (±**0.02**)
$D G^{solv}$(GC)	5.12	5.12	5.12	5.12	**8.02**
$D G^{solv}$(GC, end)	3.02	3.02	3.02	3.02	**0.24**
R^2	0.36	0.84	0.91	0.92	0.92
RMSE (training set)	0.94	0.46	0.36	0.33	0.32
RMSE (validation set)	0.92	0.37	0.30	0.38	0.41
r_i				33%	27%
r_s			43%	119%	97%

Level 2: In the next step, all the constraints introduced at level 1 were preserved, with the exception of the c_i coefficient, which was now freely optimized. Although the RMSE value of 0.46 and $R^2=0.84$ suggest a correlation (considerably better than in the previous case), the coefficient c_i becomes zero after optimization, and the present model considers again only hydrogen bonding between two strands. A plot of the calculated ΔG against the experimental values (Figure 3) shows well-differentiated groups of sequences with the same contents of CG pairs, and, due to the solvation, different end pairs.

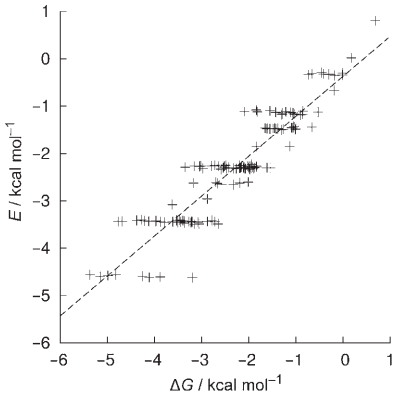

Figure 3. Estimate of ΔG based on solvation-corrected hydrogen-bonding energies plotted against experimental values.

Level 3: Stacking within one strand was now considered to be a part of duplex stabilization and, consequently, the value of the coefficient c_s was also optimized. The constraint $c_i=c_h$, used at level 1 was applied again. Table 3, column 4 shows a significantly improved correlation ($R^2=0.91$).

Level 4: At this level, all the variables (K, c_i, c_h, and c_s) were optimized independently, which resulted in the lowest value of RMSE and the best correlation (Table 3, column 5; Figure 4). Surprisingly, the RMSE of the validation set increased relative to the previous level. This might be explained by the fact that these ΔG values are not distributed evenly within the observed range, but cover the higher values only.

Let us now discuss the weight of the single energy contributions obtained at the present (most accurate) level. For this purpose, the average interaction energies of interstrand and intrastrand stacking (Table 1) weighted by the optimized values of respective coefficients were set relative to the average hydrogen-bonding energy weighted by the coefficient c_h:

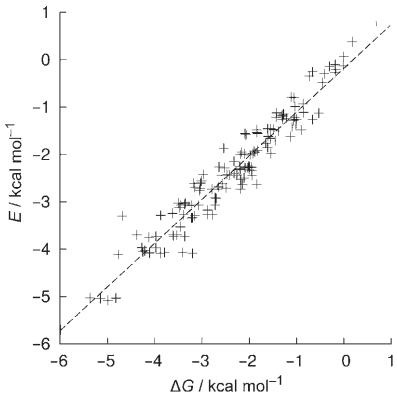

Figure 4. Estimate of ΔG composed from all calculated contributions plotted against experimental values.

$$r_i = (E^{avg}{}_i \times c_i)/(E^{avg}{}_h \times c_h) \quad (7)$$

$$r_s = (E^{avg}{}_s \times c_s)/(E^{avg}{}_h \times c_h) \quad (8)$$

The calculated r_i, which describes free-energy contribution related to interstrand stacking, amounts to 33% of the hydrogen-bonding value, although the respective interaction energy itself corresponds to only 10% of the hydrogen-bonding value (Table 1). The explanation is very complex and beyond the scope of our calculations, but two contributions could be pointed out. The first is the solvation of the basis, which weakens the free energy of hydrogen bonding. This is included in our model by introducing the $D G^{solv}$ contribution. The second one is the entropy of a complex formation. A hydrogen-bonded base pair is a rather rigid structure, whereas the stacked structures have more conformational freedom and, thus, a higher entropy term.

The calculated r_s, which describes the free-energy contribution related to intrastrand stacking, is even more important and amounts to 119%. In addition to the entropy term described in the previous paragraph, the structure of a single DNA strand is of key importance here. Because there is no generally defined geometry of a single strand of DNA that could be used as a model, we are only able to make an

assumption based on the results of the fit. The stabilizing effect of intrastrand stacking is surprisingly high, which could be attributed to a substantial loss of the stacking interaction in a single strand, due to its flexibility and the solvation of the basis.

Level 5: Contrary to previous levels, both solvation terms $D G^{solv}$ were optimized. The resulting value (Table 3, column 6) of $D G^{solv}$(CG) (8 kcal mol^{-1}) is in good agreement with a predicted range of solvation contributions (5–14 kcal mol^{-1} for the base pair in DNA and when fully solvated, respectively). The introduction of a new, freely optimized variable into the model yields reasonable results, which supports the robustness and physical significance of the model. The lower value of $D G^{solv}$(CG,end) can be attributed to the dynamic nature of DNA melting. Because the melting usually starts at the end of the oligomer, the strength of the hydrogen bonding at the end pair is more important, which is reflected in the present model by the lower destabilizing contribution for the end pairs.

Finally, we will discuss the most obvious errors arising from approximations used in our model. Firstly, we include only the interaction between DNA bases, assuming there is no sequence-dependent contribution to the overall stability originating in the backbone. Secondly, the $D G^{solv}$ term does not reflect the geometry change of a DNA strand upon dissociation. This problem is addressed by the optimization performed at level 5, and the resulting value is in good agreement with calculations. Thirdly, the intrastrand-stacking contribution is calculated as a scaled value of the interaction in a duplex. The correct procedure would be to use the difference between stacking energies in duplex and single-strand geometries. It might seem to be a rough approximation, but the only error introduced is the sequence dependency of the difference, which is expected to be low.

As well as the RI-DFT-D procedure, the cheaper DFTB-D procedure (only interaction energies were recalculated and solvation energies were kept unchanged) was also applied. The results of the fitting procedure performed at levels 1–5 are summarized in Table 4. Although the DFTB-D results are not as accurate as the RI-DFT-D ones and the differences were shown above, the correlation and RMSE results were only slightly poorer. We assign this to the systematic nature of the error, which is then handled by the fitting procedure.

Table 4. Fit of DFTB-D results to experimental unwinding energies [see Eq. (6)]. Optimized variables are shown in bold type. Standard errors of the coefficients are presented in parenthesis.

Level	1	2	3	4	5
K	**3.58** (±**1.17**)	**13.76**	**22.72** (±**0.74**)	**28.62** (±**2.02**)	**32.28** (±**1.95**)
c_h	**0.04** (±**0.01**)	**0.12**	**0.10** (±**0.003**)	**0.09** (±**0.01**)	**0.14** (±**0.01**)
c_i	$=c_h$	**0**	$=c_h$	**0.18** (±**0.03**)	**0.18** (±**0.02**)
c_s	0	0	**0.11** (±**0.003**)	**0.17** (±**0.02**)	**0.17** (±**0.02**)
$D G^{solv}$(GC)	5.12	5.12	5.12	5.12	**8.74**
$D G^{solv}$(GC, end)	−3.02	−3.02	−3.02	−3.02	−**0.77**
R^2	0.41	0.92	0.95	0.95	0.96
RMSE (training set)	1.07	0.46	0.36	0.35	0.35
RMSE (validation set)	1.22	0.53	0.39	0.40	0.50
r_i				20 %	14 %
r_s			40 %	73 %	49 %

Conclusion

1) Deep complementarity between intrastrand and interstrand stacking energies was observed over a large set (128) of DNA octamers exhibiting different sequences of DNA bases. The sum of both stacking contributions was almost constant (−99 ± 2.8 kcal mol^{-1}), whereas single components varied to a much larger extent (± 10.7 kcal mol^{-1}).
2) Close correlation between the sum of interaction and solvation energies on the one hand and DNA stability on the other was observed. All energy components, including the stacking ones, should be determined, however, as accurately as possible. The recently developed RI-DFT-D procedure was used for the calculation of hydrogen-bonding and stacking-interaction energies. The semiempirical DFTB-D method also exhibits satisfactory results; this method is very efficient and is, therefore, suitable even for large systems. The C-PCM method was used at the HF level for determining the solvation free energy.
3) The analysis of the terms contributing to the overall DNA stability obtained from our model confirmed the importance of both inter- and intrastrand stacking. Especially intrastrand stacking was found to be the major energy contribution to DNA-duplex stabilization. Our calculations also support the fact that the CG pair content in a sequence is not decisive for DNA stability. The present calculations clearly show the compensation of the strong hydrogen bonding by the favorable solvation of CG pairs.
   Themodel proposed is suitable for the prediction of the ΔG of DNA denaturation, and unlike in the case of fully empirical models, it can be extended easily to unusual DNA structures containing, for example, mismatched or unnatural base pairs.
   Incontrast to the nearest-neighbor model, in which the coefficients resulting from the fitting procedure are nonunique and have no physical meaning, coefficients in our model are directly linked to calculated variables, and, therefore, represent the relative importance of particular interactions. Our model also requires far fewer parameters (max. 6, depending on the level used), relative to the 12–42 necessary for empirical models presented in reference [3].
4) The weights of all calculated contributions were optimized to best fit the unwinding free energies over a

training set of 128 DNA octamers and were tested on a validation set of 12 octamers. In the original paper,[3] the RMSE was found by using the fully empirical nearest-neighbor model to be 0.35 kcal mol^{-1} for the validation set and 0.8 kcal mol^{-1} for the training set. We obtained the best value for the validation set, RMSE = 0.30 kcal mol^{-1}, by using the level 3 model on the RI-DFT-D data. Surprisingly, the higher level of the model yielded slightly poorer results (RMSE of 0.38 and 0.41 kcal mol^{-1} at levels 4 and 5, respectively), whereas the correlation and RMSE over the training set were better than both our lower-level models and the nearest-neighbor empirical model. The ΔG values in the validation set cover only the upper part of the range of the ΔG found in the training set, and some of them are even above the range. A comparison based on RMSE over such an unbalanced validation set is not a good measure, and we are convinced that higher-level models would yield better results if applied to a more-extended set of structures.

Acknowledgements

This work was part of the research project No. Z4 055 905 and was supported by grants from the Grant Agency of the Czech Republic (203/05/0009) and the Ministry of Education (MSMT) of the Czech Republic (Center for Biomolecules and Complex Molecular Systems, LC512).

[1] P. Jurecka, P. Hobza, *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.
[2] I. Dabkowska, H. V. Gonzalez, P. Jurecka, P. Hobza, *J. Phys. Chem. A* **2005**, *109*, 1131–1136.
[3] M. J. Doktycz, M. D. Morris, S. J. Dormady, K. L. Beattie, *J. Biol. Chem.* **1995**, *270*, 8439–8445.
[4] K. J. Breslauer, R. Frank, H. Blocker, L. A. Marky, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3746–3750.
[5] P. Jurecka, J. Sponer, J. Cerny, P. Hobza, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
[6] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, *J. Chem. Phys.* **2001**, *114*, 5149–5155.
[7] M. Orozco, F. J. Luque, *Chem. Rev.* **2001**, *101*, 203–203.
[8] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. eibel, U. C. Singh, P. K. Weiner, P. A. Kollman in *AMBER 7*, University of California, San Francisco, **2002**.
[9] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1996**, *118*, 2309–2309.
[10] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
[11] G. D. Hawkins, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem.* **1996**, *100*, 19824–19839.
[12] J. Srinivasan, M. W. Trevathan, P. Beroza, D. A. Case, *Theor. Chem. Acc.* **1999**, *101*, 426–434.
[13] P. Jurecka, J. Cerny, P. Hobza, D. R. Salahub, *J. Comput. Chem.* **2006**, in press.
[14] K. Eichkorn, O. Treutler, H. Ohm, M. Haser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283–289.
[15] R. Ahlrichs, M. Bar, M. Haser, H. Horn, C. Kolmel, *Chem. Phys. Lett.* **1989**, *162*, 165–169.
[16] V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
[17] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Wallingford CT, **2004**.
[18] J. Petruska, M. F. Goodman, *J. Biol. Chem.* **1995**, *270*, 746–750.
[19] P. Hobza, J. Sponer, *Chem. Rev.* **1999**, *99*, 3247–3276.

Received: August 2, 2006
Published online: December 21, 2006

