Intra-strand stacking interactions in B-DNA crystals characterized by post-SCF quantum chemistry computations†

Piotr Cysewski*

Received (in Montpellier, France) 11th May 2009, Accepted 10th July 2009
First published as an Advance Article on the web 4th August 2009
DOI: 10.1039/b909327c

The distributions of intermolecular interaction energies (IIE) of all possible 5′-X/Y-3′ (X, Y = A, G, T, C) stacked pairs in conformations exactly matching to intra-strand orientations occurring in crystallographic B-DNA double strands were characterized by quantum chemistry calculations on MP2/aug-cc-pvDZ level. The essential feature of this approach is sampling over meaningful conformations and getting statistically significant data on the energetics of stacking interactions in B-DNA crystals. Based on the most frequently occurring IIE values (in kcal mol^{-1}) the following order of the intra-strand stacking interactions in B-DNA solids is concluded: G/A -10.2 ± 1.1 > A/G $-9.2 \pm 1.0 \approx$ G/C -8.7 ± 1.2 > A/A -8.3 ± 1.0 > A/T -7.7 ± 0.9 > C/G -7.2 ± 0.9 > T/G $-6.5 \pm 1.3 \approx$ A/C $-6.2 \pm 0.8 \approx$ G/T -6.2 ± 1.3 > T/A -5.6 ± 0.9 > G/G -5.1 ± 1.1 > T/C -4.9 ± 0.7 > C/T -4.8 ± 1.1 > T/T -4.7 ± 0.5 > C/A -3.8 ± 0.8 > C/C -1.0 ± 1.0 (standard deviations are provided in parenthesis). Sequence-related structural diversities and SCF and correlation contributions to the stacking interactions are also discussed. The most representative structures of stacked dimers were found based on a clustering protocol.

Introduction

Contemporary instrumental methods provide direct and valuable insight into the structure of bio-molecules such as proteins, nucleic acids or their diverse complexes. Atomic-resolution X-ray diffraction patterns and NMR data provide essential information on all kinds of contacts occurring within realistic environments.[1] However, although these techniques reveal structures the energetic characteristics are not directly available. Fortunately, quantum chemistry tools, as complementary to experiments, provide additional information if adequate level of theory is applied. Great efforts have been made for characterization of intermolecular interactions stabilizing the nucleic acid double helix.[2–6] Among many contributions, the non-covalent interactions of nucleic acid bases are of particular importance,[2] as essential factors stabilizing the helical structure of DNA.[6] Since the non-covalent interactions are governed mainly by London-dispersion contribution, their theoretical description is computationally demanding and only the most accurate approaches lead to adequate results.[6] Although the X-ray diffraction images of DNA crystals[1] provide precise structural information on atomic resolution of stacked pairs in B-DNA, the direct utilization of available Cartesian coordinates is not recommended for *ab initio* calculations.[3,7] Nevertheless, the base pair and base step parameters may be successfully used for accurate definition of nucleobases arrangements in B-DNA conformations. This procedure was nowadays successfully applied for polymorphism-related heterogeneities of guanine stacking in B- and A-DNA forms,[8] characteristics of inter- and intra-strand stacking interactions in d(CpG) and d(GpC) steps found in B-DNA, A-DNA and Z-DNA crystals[9] and description of energetic heterogeneities in canonical and oxidized central guanine triads of B-DNA telomeric fragments.[10] The main advantage of such a procedure is the direct use of the conformations occurring in nature[8–11] instead of model ones typically applied in theoretical studies.[12–14] Besides, the expression of structural information in commonly accepted terms[1] and application of advanced computational methods makes unique linkage between quantum chemistry and structural molecular biology. Thus, the aim of this paper is to analyze the distributions of intermolecular interaction energies (IIE) of all possible intra-strand stacked pairs in conformations exactly matching to ones occurring in crystallographic B-DNA double helices. The crucial feature of this approach is the way of sampling over meaningful conformations and getting statistically significant data on the energetics of stacking interactions in B-DNA crystals.

Methods

The base pair (*shear*, *stretch*, *stagger*, *buckle*, *propeller* and *opening*) and base step parameters (*shift*, *slide*, *rise*, *tilt*, *roll* and *twist*) taken from the Nucleic Acid Database (NDB)[1] were used for preparation of all possible X/Y (where X, Y = {A, G, C, T}) pairs of nucleobases in conformation matching B-DNA crystals. In this paper X/Y means 5′-X/Y-3′ intra-strand stacking occurring either in I or II strand. Obviously 5′-X/Y-3′ is structurally and energetically different than 5′-Y/X-3′. Hence, there are 16 possible combinations of two

Department of Physical Chemistry, Collegium Medicum, Nicolaus Copernicus University, Kurpińskiego 5, 85-950 Bydgoszcz, Poland. E-mail: piotr.cysewski@cm.umk.pl
† Electronic supplementary information (ESI) available: Fig. S1–S13 and Table S1. See DOI: 10.1039/b909327c

intra-strand stacked pairs. Since technical details were presented elsewhere[8–10] here only a brief summary is provided. Among all available structures in the NDB only those were taken into account, which were related to native double helices without any ligands, proteins, with no mismatches or any modifications of nucleobases, sugar or phosphate moieties. For getting as many structures as possible no predefined selection of B-DNA files was applied. The X3DNA program[1] with a modified internal library was used for preparation of pairs of given sequence based on all 18 parameters extracted from PDB files. The nucleobase coordinates defined in the 3XDNA local library (without hydrogen atoms) were modified and replaced with ones optimized at MP2/aug-cc-pVDZ (aDZ) level in C_s symmetry. Then all atoms were removed from the two base pair steps except those belonging to intra-strand stacked pairs of nucleobases. In such a way the sugar-phosphate backbones were simplified just by hydrogen atoms. The applied procedure leads to pairs of stacked nucleobases exactly matching to ones that are present in B-DNA crystals. However, strict planarity of nucleobases is assumed including amino side-groups. Although Hobza[2] demonstrated significant out-of-pane deformation of amino-side groups it is assumed that this effect does not affect seriously stacking interactions. Single point energy calculations at DF-MP2/aDZ level of theory were performed for each stacked pair and resulting intermolecular interaction energies (IIE) of stacked complexes were estimated. Due to the fact that the dominant interaction-energy contribution to stacking interactions comes from London dispersion energy the application of highly advanced quantum chemistry calculations is required to guarantee sufficient accuracy.[2–7] Although it was emphasized many times in the literature[2–7] that CCSD(T) procedure and extended basis sets are indispensable for proper assessment of stacking energetics there were also applied much less demanding procedures.[2,12] The fortunate cancellation of errors makes the computations of stacking interactions feasible.[3] For example it is well known that the MP2 method overestimates the attraction between aromatic rings significantly compared with more reliable CCSD(T) calculation.[6] This is even more pronounced for extended basis sets including CBS procedure (complete basis sets extrapolation).[2] However, coupled cluster corrections are usually positive and very often the MP2 along with medium basis sets give quite reasonable estimation of stacking interactions. This is presented in Fig. 1 for data coming from a so-called benchmark database.[4] Although all intra- and inter-strand stacking are overestimated by MP2/aDZ method compared to CCSD(T), there are still quite acceptable linear relationships. This paper deals with statistically significant distributions of nucleobase pairs and, as will be presented below, the resulting standard deviations estimated for analyzed populations usually exceed the error introduced by MP2/aDZ method. Thus, the MP2/aDZ method seems to be a reasonable compromise between accuracy and efforts. Extension of computations beyond this level seems to be not necessary. The counterpoise correction for BSSE error[14] was included in all single point calculations of stacked complexes. This is very important since the magnitude of this error is comparable to values of pair stabilization energy. This was demonstrated in

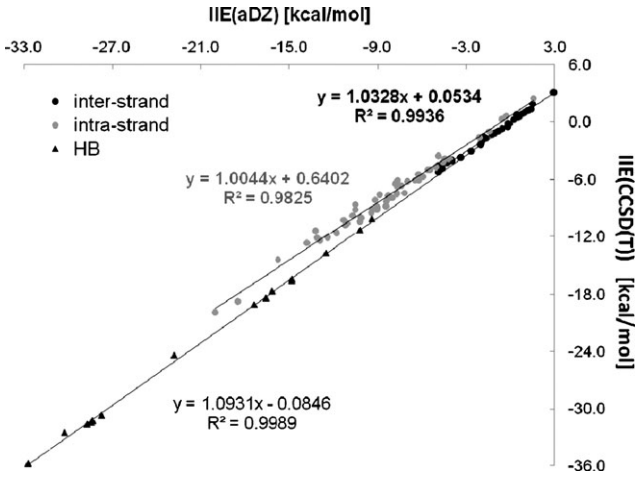


Fig. 1 The correlation between hydrogen bonding, inter- and intra-stacking interaction energies estimated on MP2/aDZ and CCSD(T) levels of theory. All values were taken from ref. 4.

Fig. 2 collecting all obtained values of stacking interactions. In all calculations of intermolecular interaction energies (IIE) the MolPro package[15] was used and density fitting procedures[7] were applied for MP2 computations.

Results and discussion

The IIE values were estimated for all possible nucleobases pairs in stacked orientations. There are 16 unique sequences, namely A/A (152), A/G (69), A/C (59), A/T (62), G/A (101), G/G (106), G/C (112), G/T (59), C/A (61), C/G (132), C/C (108), C/T (69), T/A (64), T/G (61), T/C (102), T/T (150). In parenthesis the number of pairs that were used in this project are provided. The stacked pairs belonging to strands I and II are structurally and energetically indistinguishable and both were included. For example the d(ApG) steps are the source of both 5′-A/G-3′ and 5′-C/T-3′ stacked pairs since base step parameters are the same for both of them. Three aspects were analyzed in detail, namely IIE distributions, structural diversities and correlation between conformation and stacking interactions of stacked pairs present in B-DNA crystals. Besides, most representative structures (MRS) were identified for each of the stacked sequences. The structure comparison was done by utilization of normalized values of root mean square deviation (NRMSD) defined in the following formula:

$$\mathrm{NRMSD}(i,j) = \sqrt{\sum_{k=1}^{6} (X_k(i) - X_k(j))^2}$$

where X stands for normalized values of base step parameters (*i.e.* 1-*shift*, 2-*slide*, 3-*rise*, 4-*roll*, 5-*tilt*, 6-*twist* denoted by index k) of i-th and j-th stacked pairs. Normalization was done for preserving of all values within the interval from 0 to 1 for each variable and for all of analyzed pairs. The values of NRMSD are the measure of similarities of stacked nucleobases conformations. All NRMSD distributions were presented as supplementary material (ESI†) in Fig. S1–S4 and below only representative plots are discussed (see Fig. 3 and 4). Furthermore, the most representative structures were saught

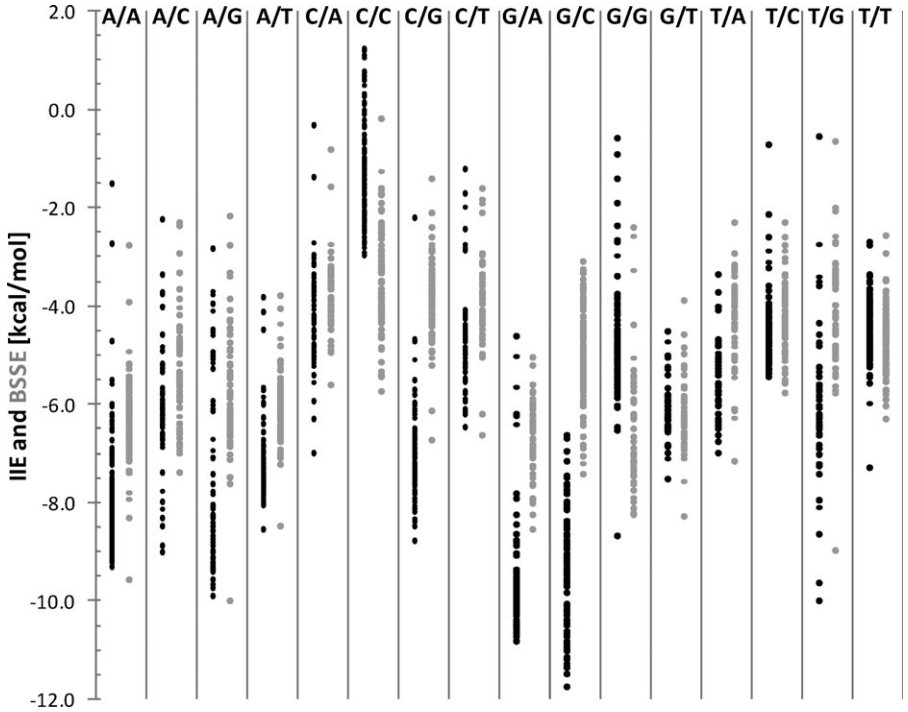


Fig. 2 The values of intermolecular interaction energies (IIE-black dots) and BSSE (gray dots) of all possible 5′-X/Y-3′ ({X, Y} = {A, G, T, C}) pairs of nucleobases in conformations found in B-DNA crystals.[1]

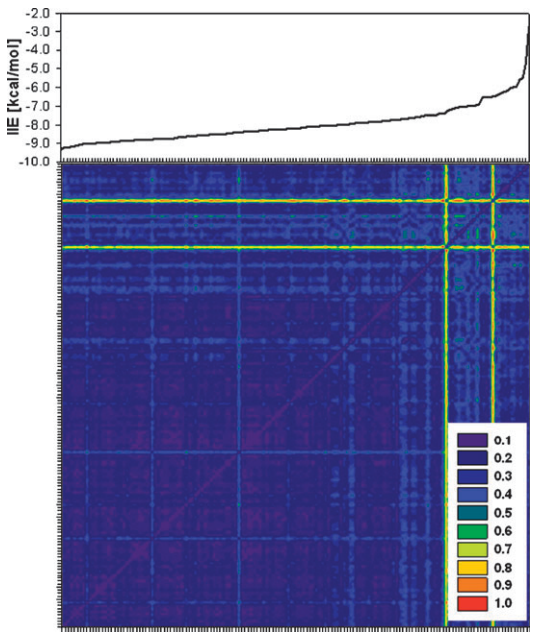


Fig. 3 The distributions of NRMSD values (RMSD of normalized values of base step parameters: *shift*, *slide*, *rise*, *roll*, *tilt*, *twist*) characterizing structural diversity of 5′-A/A-3′ stacked dimers in conformations occurring in B-DNA crystals.[1] The pairs are denoted by tick marks on edges of the plot and were ordered with increasing values of stacking energy (IIE in kcal mol^{-1}). The color scale represents actual values of NRMSD.

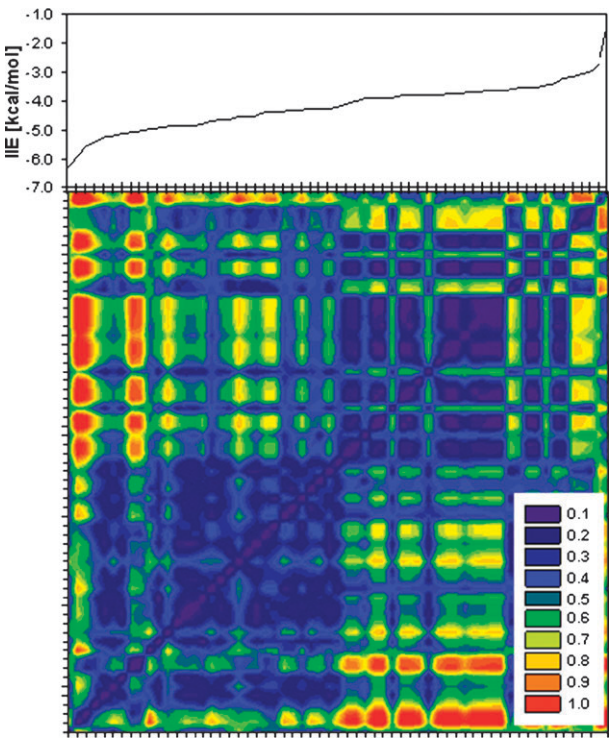


Fig. 4 The distributions of NRMSD values characterizing structural diversity of 5′-C/A-3′ pairs in conformations occurring in B-DNA crystals.[1]

by calculation of the mean NRMSD value for each pair with respect of all other stacked complexes of particular sequence. This parameter has the meaning of closeness of a particular pair to all the other dimers of the same type. The stacked pairs, that are characterized by smallest values of mean NRMSD are termed the most representative structures (MRS). The

correlation between values of mean NRMSD and IIE are presented as contour plots in Fig. S10–S13 (ESI†).

Distribution of IIE in B-DNA crystals

The IIE distributions characterizing all nucleobases sequences are presented in Fig. 2. The energetic heterogeneities of stacked pairs are clearly visible since one can find structures with high stabilization interactions as well as ones with no attraction or even destabilization contribution of stacking to the total energy of the polynucleotide chain. This is of course related to the variations of local conformations of the B-DNA double helix. Among all pairs considered here, the strongest interactions are found for G/C couples in B-DNA denoted as bd0064,[1] for which IIE = −11.7 kcal mol^{-1}. Significant sequence dependence of IIE is observed since the most stable C/C pair occurring in bd0065 is stabilized by only −3.0 kcal mol^{-1}. For the remainder of stacked nucleobases the highest attraction occurs in between these values. However, these extreme IIE values are relatively rare, and from the statistical point of view, they are not the most important ones due to their small contribution to the whole population of stacked pairs found in B-DNA crystals. Instead, the most frequently occurring (MFO) intermolecular interaction energies (IIE^{MFO}) are a better representation of intra-strand stacking. For all analyzed X/Y sequences the difference between smallest and highest IIE values of the most frequently occurring pairs does not exceed 1.4 kcal mol^{-1}. The highest divergence is observed for G/C and C/G pairs (1.4 kcal mol^{-1}) and the smallest for T/T (0.6 kcal mol^{-1}). Thus, despite of the significant structural heterogeneity of nucleic acid bases in stacked conformations the energies of most frequently occurring stacked pairs of particular context are very similar. This conclusion is of great importance from the perspective of DNA stabilization forces. For a broad range of values of structural parameters defining pairs conformations the similar intermolecular interactions guarantee flexibility and energetic stability of the B-DNA double helix. If strong attractions of stacking interaction were present only for a narrow window of conformations, the B-DNA would be very stiff, unstable and very sensitive to geometry fluctuations. The observed great elasticity of DNA is due to the flexibility and low-diversity of stacking interactions.[2–7] Such property of stacking interactions is also expected for other polymorphic forms of DNA.[8] Based on the most frequently occurring IIE values the following energetic succession of stacked nucleobases in the whole population of analyzed B-DNA crystals can be proposed: **G/A** (−10.2 ± 1.1) > ($p < 0.001$) **A/G** (−9.2 ± 1.0) ≈ ($p = 0.055$) **G/C** (−8.7 ± 1.2) > ($p < 0.001$) **A/A** (−8.3 ± 1.0) > ($p < 0.001$) **A/T** (−7.7 ± 0.9) > ($p < 0.005$) **C/G** (−7.2 ± 0.9) > ($p < 0.001$) **T/G** (−6.5 ± 1.3) ≈ ($p = 0.902$) **A/C** (−6.2 ± 0.8) ≈ ($p = 0.882$) **G/T** (−6.2 ± 1.3) > ($p = 0.009$) > **T/A** (−5.6 ± 0.9) > ($p = 0.008$) **G/G** (−5.1 ± 1.1) > ($p = 0.002$) **T/C** (−4.9 ± 0.7) > ($p = 0.002$ **C/T** (−4.8 ± 1.1) > ($p < 0.001$) **T/T** (−4.7 ± 0.5) > ($p < 0.001$) **C/A** (−3.8 ± 0.8) > ($p < 0.001$) **C/C** −1.0 ± 1.0. The values of standard deviations are provided in parenthesis. Besides, next to the comparison quantifier there are values of the probabilities that the difference in the medians of two successful distributions are greater than would be expected by chance. The statistical significance of the difference between two distributions was analyzed based on the non-parametrical Mann–Whitney rank sum test.[16] The majority of IIE distributions are statistically distinct. However, there are distributions, which are almost identical. For example IIE distributions of A/G and G/C are statistically indistinguishable with 95% confidence. Similarly, the IIE populations of A/C, G/T and T/G are equivalent. The most frequently occurring stacking energies of G/A pairs are characterized by strongest attractions in B-DNA. On the contrary the lowest value of IIE^{MFO} corresponds to C/C pairs. This proves significant sequence dependence of stacking in B-DNA crystals. Furthermore, the context-dependence is also manifested for all analyzed pairs except from guanine stacking with thymine. This is the only case where inversion of 5′- and 3′-sides does not affect the intermolecular interactions. In all other situations the increase in IIE^{MFO} is caused by purine presence on 5′-side which is especially pronounced for guanine. The stacking interactions in homo-nucleobases pairs are stronger for purines than for pyrimidines. Most of distributions presented in Fig. 2 are non-normal having an asymmetric tail at the higher IIE side. This may indicate that there are conformations disfavouring stacking. The observed diversities of base pair and bases step parameters[8] may be the primary source of such IIE distributions. Alternatively, the potential errors introduced during the refinement process of X-ray diffraction patterns may also be the source of such asymmetry. Often nucleobases pairs characterized by high IIE values come from the same files (*i.e.* bdj081, bd0022, bd0082, bd0089 and bd0059—measured with resolutions equal to 1.85, 2.60, 2.00, 3.10 and 2.50 Å, respectively). At this point it is worth emphasizing that conclusions related to energetic and local structure of B-DNA must be treated with caution if they are based only on one particular PDB file or a small sample of structures. Fortunately, the percentage of such suspicious files is very small and does not distract the main features of IIE distributions. The analysis of statistically meaningful data presented here has the obvious advantage of being less affected by data inaccuracies.

Structural diversity of stacked pairs

The clustering of structures can be performed by analysis of one-to-one relations of geometries and energies. The similar values of RMDS indicate structural correspondence. As it was mentioned beforehand the NRMSD values were used here for description of structural diversities of stacked pairs. Although NRMSD values do not provide information about the source of eventual structural dissimilarities they still are a useful measure of variations of base step parameters. All 16 distributions are provided in ESI† (see Fig. S1–S4) and here only two were selected as most representative types. In Fig. 3 the clusters of A/A pairs are presented. All points on the diagonal correspond to zero values of NRMSD since structures are compared to themselves. Of course distributions are symmetrical and points above and below the diagonal have the same meaning. Fig. 3 presents a quite extreme case since

most pairs are structurally very similar. Also more than 90% of corresponding stacking energies differ from IIE^{MFO} by less than 2 kcal mol^{-1}. Thus, in this case the majority of stacked pairs exhibit great structural and energetic similarities. There are few exclusions for which NRMSD increase without significant alteration of stacking interactions but most of pairs fall into one broad class of structures. Small variations of NRMSD are observed also for such pairs as T/T, AT or CG.

They either adopt conformations within a narrow range of base step parameters or similar NRMSD values are obtained by mutual compensation of these parameters. On the contrary, the second extreme corresponds to C/A pairs as is illustrated in Fig. 4. It is evident that there are many dissimilar structures with comparable stacking interactions. This strongly suggests that there is not just one conformation, which is preferentially adopted in crystals by stacked C–G and A–T pairs constituting the d(CpA) step. Not only the mean structure but also the so-called "ideal" B-DNA form should be used with great care as it is not dominant in crystals. Even finding of the most representative structure is problematic in this case. This means that many quite different base–base conformations may be energetically equivalent. The same feature is typical also for T/A, C/A (or T/G), G/G (or C/C) and G/C pairs. The remainder of base sequences behave in between these two extremes. The origin of NRMSD variations may be inferred from distributions of base step parameters. In Fig. 5 there are presented variations of *slide* values corresponding to those structures, which exhibit greatest similarities. The values of *slide* were presented only for those structures, which are within a standard deviation from the minimum value of mean NRMSD. The distributions of the remainder of the base step parameters are collected in ESI† in Fig. S5–S9, respectively. As it can be inferred from Fig. 5 distributions of *slide* are sequence-dependent. Not only the number and span of most representative values is context-related but also the median values are distinct for different sequences. *Slide* representing the mutual displacement of two successive base pairs along the longer molecular axis is a base step parameter, which is the most sensitive to the sequence of stacked base pairs. It is worth emphasizing that variations of the remainder of base step parameters is much smaller and the median values corresponding to all X/Y sequences are within standard deviations for all base step parameters (see Fig. S5–S9 in ESI†). The *slide* parameter corresponds to the highest structural diversities of stacked nucleobases in B-DNA crystals and it can be considered as the main source of NRMDS heterogeneities.

Diversity of stacking interactions in B-DNA crystals

The structure–energy relationships of stacked pairs is a very complicated and non-linear function. This very important subject of structural biology was studied extremely extensively.[19–23] Unfortunately, till now no acceptable solution was provided. One of the reason of this situation is the very complex nature of such a function. As an illustration of this feature two contour plots are provided in Fig. 6 and 7. They present two extreme examples of correlations between values of IIE and mean NRMSD. The color spectrum represents percentages of structures characterized by structural and energetic similarities. The whole interval of IIE and mean NRMSD values was divided into ten equal intervals and the percentage of such structures were estimated and plotted as smoothed contours. Details for all of X/Y stacked complexes (see Fig. S10–S13) are included in ESI.† Data presented in Fig. 6 are in line with the conclusions drawn above. Most A/A structures belong to one class with significant energetic and structural similarities. This is represented by the high and narrow peak in Fig. 6. A/A dimers occasionally occur for which higher structural fluctuations are observed but the number of such clusters is marginal for the whole population. This is not the case for two stacked cytosine molecules. As it is presented in Fig. 7 there are many conformations which are

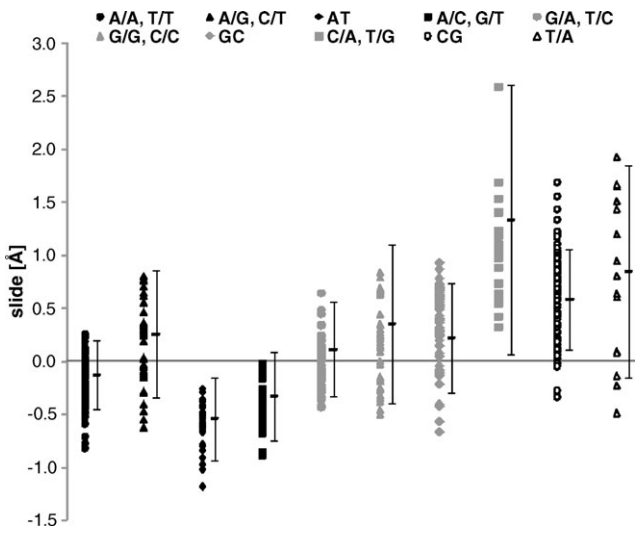


Fig. 5 The distributions of *slide* values characterizing most representative structures unique stacked pairs. The mean value estimated for all stacked pairs along with corresponding standard deviation bars are placed next to each distribution.

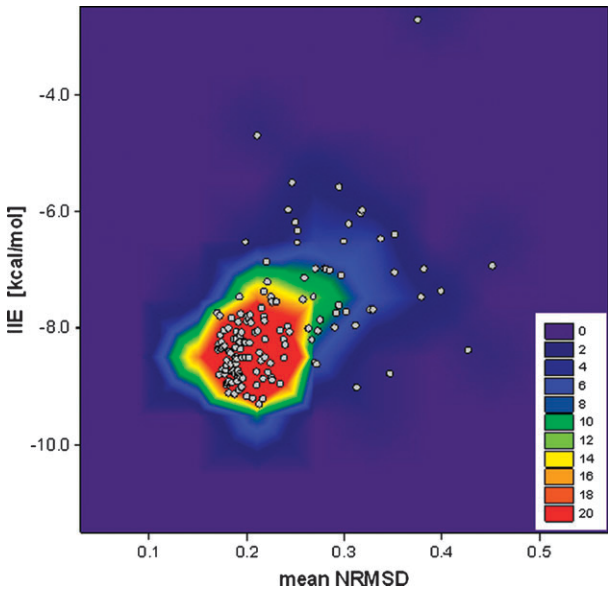


Fig. 6 The correlation between mean values of NRMSD and IIE of 5′-A/A-3′ pairs in conformations occurring in B-DNA crystals.[1] The color spectrum represent population percentages.

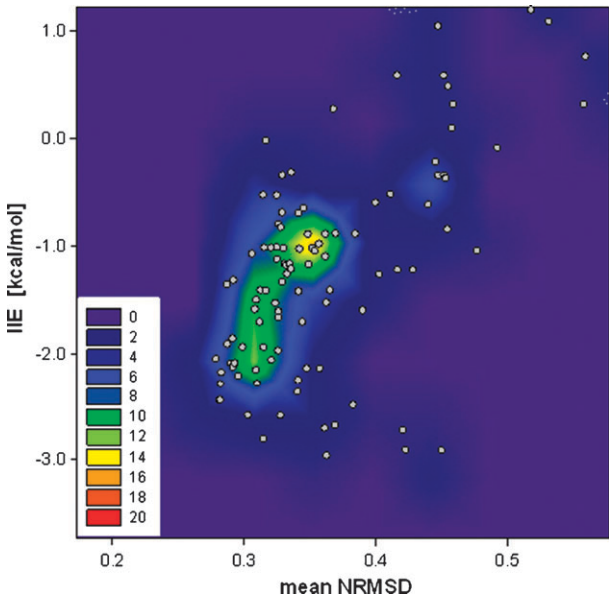


Fig. 7 The correlation between mean values of NRMSD and IIE of 5′-C/C-3′ pairs in conformations occurring in B-DNA crystals.[1]

distinct. The observed diversity of stacked structures originates from broad range of base step parameters defining structures of C/C pairs. Furthermore, there is an interesting bimodal feature of IIE distribution for these stacked pairs. As is demonstrated in Fig. 7 there are several structures, which have similar mean NRMSD values but differ in stacking interactions demonstrated by two almost separated peaks on provided contour plots. There is quite simple origin of this fact. After inspection of the distributions of base step parameters the only bimodal character is found on *shift* and *slide* histogram. There are about 41% of C/C pairs with negative values of *shift*. Interestingly, for A/A pairs exactly half of all pairs adopts negative and positive values of this parameter. Thus, in case of C/C dimers this is *shift*, which is the source of observed bimodal character of IIE distributions. The mean values of *shift* estimated for populations with negative and positive signs are equal to -0.57 ± 0.38 Å and $+0.48 \pm 0.36$ Å, respectively. The analogical values for A/A pairs are equal to -0.32 ± 0.31 Å and $+0.26 \pm 0.24$ Å, respectively. Hence, not only proportions are different for analyzed cases but also the mean values are not equivalent. *Shift* has a meaning of displacement of two successful base pairs along the shorter molecular axis with positive values towards the major groove. Then, it seems that C/C in B-DNA crystals can be packed in two distinct ways affecting major and minor grooves. Since G/G and C/C pairs share the same set of bases step parameters, the above conclusion should be extended for d(GpG) steps (indeed it is indicated for G/G pairs in Fig. S11 in ESI†). Additionally, 88% of those C/C pairs with stabilization energy higher than -2.0 kcal mol^{-1} adopt negative values of *slide*. Among those structures, which are characterized by repulsive stacking interactions about 93% has positive *slide*. Furthermore, significantly higher values of *tilt* and *twist* are related to those pairs. This combination of base step parameters usually leads to very small overlapping of pyrimidine rings what results in repulsion between stacked cytosine molecules.

One can consider another way of description of structural diversity of stacking interactions in B-DNA crystals. The lowest values of mean NRMSD for each of the pair sequences represent the overall similarity of particular stacked clusters. According to plots presented in Fig. 6 and 7 it is reasonable to expect much lower values of mean NRMSD for A/A (and T/T) than for C/C (and G/G) pairs. Indeed data collected in Table 1 confirm such expectation. In this table names of structures which exhibit greatest similarity to all others of the same sequence are provided. The value of minimal mean NRMSD is in correlation with previously discussed clustering of structures. Thus, for each set of analyzed stacked pairs one particular structure is eventually to be considered as the most representative one. This is justified also by the fact that the stabilization energies of such pairs are very close to most frequently occurring IIE values and always are within the standard deviations to IIE^{MFO}. The full characteristics of base step parameters defining conformations of most frequently occurring structures is provided in Table 2. In all cases, the values of these parameters are within a standard deviation from the mean, values which are also included in the Table 2.

Decomposition of IIE into electrostatic and dispersion contributions

It was demonstrated[17,18,23] that for a broad range of monomer orientations the electrostatic and dispersion contributions are dominant terms in the base pair stabilization energy.[24–26] The full energy decomposition is a very demanding task because the variational–perturbational approach is required for proper inclusion of electron correlation. Besides, extended basis sets are indispensable. Unfortunately, the application of this procedure to the set of analyzed pairs is impractical. However, simplified analysis is possible, which accounts only the amount of SCF and electron correlation interaction energies. It is well known that hydrogen-bonded complexes are stabilized mainly by the electrostatic interactions and one-electron approximation is generally sufficient for description of stabilization of such complexes because the electron correlation has usually a small contribution in such cases. On the contrary for stacking interactions the electron correlation is essential. It is interesting to see if the mutual compensation of these two contributions takes place for analyzed pairs. The results of such limited decomposition are presented in Fig. 8, where the correlation between median values of SFC and electron correlation distributions are presented. The total IIE values are simply the sum of these two components. It is evident that for most of stacked complexes the SCF term is positive. The source of this property is the relatively high electrostatic repulsion which is especially manifested in the following populations of stacked pairs: G/G ($+8.1$), C/C ($+5.8$), A/A ($+5.8$), A/C ($+5.1$) and G/T ($+5.0$), where in parenthesis the median values of $IIE^{HF(aDZ)}$ distributions are provided (in kcal mol^{-1}). Interestingly, in case of C/G (-1.2) and G/C (-0.5) electrostatic attractions occur between stacked monomers. This is a rather unexpected feature since it is commonly accepted that the dispersion is the main source of

Table 1 The energetic characteristics (in kcal mol^{-1}) of intra-strand stacking in conformations of two nucleobases occurring in B-DNA crystals. The number in parenthesis next to the NDB file name represents the position of a given pair in the oligonucleotide chain

Pair	IIE^{MFO}	μ^a	$\Delta\Delta G^{sol}$	NDB file	NRMSD^b	IIE^{MRS}
A/A	−8.3 ± 1.0	4.2 ± 0.1	8.9 ± 0.4	bdl001 (5)	0.17	−7.8
A/G	−9.2 ± 1.0	4.6 ± 0.2	10.3 ± 0.4	bdj060 (3)	0.25	−8.1
A/T	−7.7 ± 0.9	5.7 ± 0.1	9.4 ± 0.4	bd0041 (6)	0.17	−7.6
A/C	−6.2 ± 0.8	8.2 ± 0.1	7.8 ± 0.2	bd0001 (11)	0.19	−6.3
G/A	−10.2 ± 1.1	7.4 ± 0.2	11.8 ± 0.4	bdl001 (8)	0.18	−10.3
G/G	−5.1 ± 1.1	12.2 ± 0.3	7.6 ± 1.1	bd0039 (1)	0.27	−4.9
G/T	−6.2 ± 1.3	9.1 ± 0.3	9.2 ± 0.5	bd0001 (11)	0.19	−6.2
G/C	−8.7 ± 1.2	6.9 ± 0.6	12.7 ± 0.7	bdl007 (2)	0.27	−9.4
C/A	−3.8 ± 0.8	6.9 ± 0.4	8.1 ± 1.3	bd0090 (10)	0.34	−4.2
C/G	−7.2 ± 0.9	1.2 ± 0.6	11.5 ± 0.6	bd0054 (11)	0.24	−7.0
C/T	−4.8 ± 1.1	7.8 ± 0.3	9.0 ± 0.7	bdj060 (3)	0.25	−5.9
C/C	−1.0 ± 1.0	11.9 ± 0.2	4.7 ± 0.9	bd0039 (1)	0.28	−2.0
T/A	−5.6 ± 0.9	6.5 ± 0.1	9.5 ± 0.5	bd0077 (5)	0.33	−6.0
T/G	−6.5 ± 1.3	5.0 ± 0.7	10.8 ± 1.3	bd0090 (10)	0.34	−7.0
T/T	−4.7 ± 0.5	8.0 ± 1.0	8.7 ± 0.6	bdl001 (5)	0.16	−4.4
T/C	−4.9 ± 0.7	10.5 ± 0.1	8.1 ± 0.5	bdl001 (8)	0.18	−4.8

[a] Mean values of dipole moment (in Debye) in the gas phase. [b] The minimal value of mean NRMSD.

Table 2 The mean vales along with standard deviations of base step parameters characterizing pairs of stacked nucleobases found in B-DNA crystals. In parenthesis the values corresponding to most representative structures are provided, for which the names are given in Table 1. The second column gives the percentage of the most representative structures in the whole populations of stacked pairs

	%	*Shift*	*Slide*	*Rise*	*Tilt*	*Roll*	*Twist*
A/A, T/T	71	0.0 ± 0.4 (0.1)	−0.1 ± 0.3 (−0.3)	3.3 ± 0.2 (3.3)	0.2 ± 3.3 (−0.7)	0.0 ± 4.2 (0.9)	35.9 ± 3.7 (35.4)
A/G, C/T	61	0.0 ± 0.6 (−0.2)	0.3 ± 0.6 (0.3)	3.2 ± 0.3 (3.2)	1.0 ± 4.3 (2.5)	3.6 ± 4.0 (3.5)	31.4 ± 5.6 (28.6)
AT	73	0.0 ± 0.6 (0.0)	−0.5 ± 0.4 (−0.6)	3.3 ± 0.2 (3.3)	−0.5 ± 3.9 (−0.2)	−0.7 ± 3.2 (−0.9)	33.8 ± 4.0 (34.2)
A/C, G/T	59	0.0 ± 0.6 (0.0)	−0.3 ± 0.4 (−0.4)	3.3 ± 0.2 (3.3)	0.3 ± 3.3 (0.2)	0.8 ± 3.6 (1.4)	33.7 ± 4.2 (33.8)
G/A, T/C	63	0.0 ± 0.6 (0.1)	0.1 ± 0.4 (0.0)	3.4 ± 0.2 (3.4)	−0.3 ± 3.3 (0.3)	1.0 ± 3.7 (−0.1)	38.7 ± 3.7 (38.4)
G/G, C/C	44	0.0 ± 0.6 (0.4)	0.4 ± 0.7 (0.1)	3.3 ± 0.3 (3.3)	−0.5 ± 5.0 (−0.7)	3.5 ± 4.5 (5.4)	33.2 ± 5.6 (31.6)
GC	48	−0.1 ± 0.8 (0.3)	0.2 ± 0.5 (0.5)	3.4 ± 0.2 (3.5)	−0.5 ± 2.8 (2.6)	−3.0 ± 5.9 (−3.4)	37.9 ± 4.5 (38.0)
C/A, T/G	41	0.0 ± 0.4 (0.0)	1.3 ± 1.3 (1.7)	3.3 ± 0.2 (3.3)	0.0 ± 2.3 (1.5)	0.8 ± 6.6 (−0.3)	39.8 ± 9.2 (39.2)
CG	86	0.0 ± 0.6 (−0.2)	0.6 ± 0.5 (0.4)	3.3 ± 0.3 (3.3)	0.2 ± 3.1 (−2.2)	4.5 ± 5.1 (6.9)	32.7 ± 5.1 (32.0)
T/A	25	0.0 ± 0.6 (−0.4)	0.8 ± 1.0 (1.4)	3.3 ± 0.3 (3.4)	−0.6 ± 3.2 (0.5)	0.4 ± 5.8 (−3.5)	38.8 ± 7.0 (38.4)

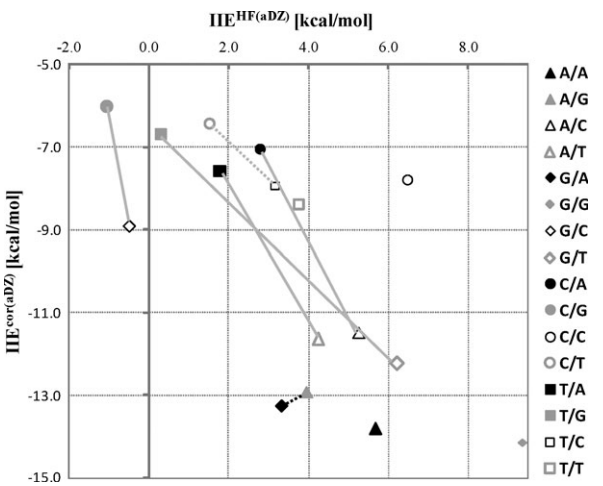
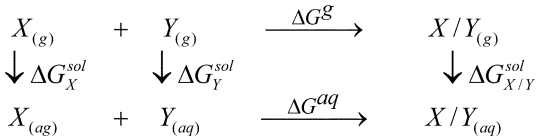


Fig. 8 The correlation between SCF (IIE^{HF}(aDZ)) and (IIE^{corr}(aDZ)) electron correlation contributions to the total inter-molecular interaction energies of stacked pairs in conformations corresponding to B-DNA crystals. The plotted values correspond to medians of IIE^{HF}(aDZ) and IIE^{corr}(aDZ) distributions. Lines connects pairs of the same monomers oriented in two opposite contexts.

stacking stabilisation. However, for these three sequences the stabilization of stacked complexes may also originate from the electrostatic interactions. As it is demonstrated in Fig. 8 the electron correlation corresponding to all sequences of stacked pairs is always negative and exceeds the electrostatic repulsion. The highest amount of dispersion is expected for A/A (−13.7), G/A (−13.6), G/G (−13.3) and A/G (−12.8), where values in parentheses correspond to IIE$^{corr(aDZ)}$ medians (in kcal mol^{-1}). Purines have much stronger influence on the stacking then pyrimidines and determine the amount both of electron correlation and total IIE values. For homo-nucleobases dimers formed between two purines the dispersion is much stronger than for homo-pyrimidine pairs. Interestingly, the corresponding magnitude of SCF energy is comparable, which explains why purines form significantly more stable stacked pairs. If purine and pyrimidine are involved in stacking the presence of the former at 5′-side always leads to decrease of IIE$^{corr(aDZ)}$ values and increase of the IIE$^{HF(aDZ)}$ term. The first trend is dominant and consequently A/C, G/C, A/T and G/T pairs are more energetically favourable then C/A, C/G, T/A and G/T, respectively. There is a simple geometric rationale of this fact since purine molecules introduce more delocalised π-electron densities and polar side groups to a stacked pair.

Solvent influence on base–base stacking interactions

Finally, the influence of the electrostatic field on the stabilization energy of stacked pairs is analyzed. Let consider Scheme 1 describing formation of stacked complex (X/Y) from



Scheme 1

monomers (X and Y) in the gas phase (g) and in aqueous solution (aq).

The relative solvation Gibbs free energy can be directly estimated from above thermodynamic cycle *via* the following formula:

$$\Delta\Delta G^{\mathrm{sol}} = \Delta G^{\mathrm{aq}} - \Delta G^{\mathrm{g}} = \Delta G^{\mathrm{sol}}_{X/Y} - \Delta G^{\mathrm{sol}}_{X} - \Delta G^{\mathrm{sol}}_{Y}$$

The above equation defines the relative stability of stacked 5′-X/Y-3′ pair in aqueous solution with respect to the gas phase. It is to be considered as a measure of hydration of stacked pair, and the higher the value of $\Delta\Delta G^{\mathrm{sol}}$, the less stable the stacked complex in solution. Since precise estimation of ΔG^{g} is not a trivial task even for isolated molecules[27] the relative value $\Delta\Delta G^{\mathrm{sol}}$ seems to be more accurate due to expected cancellation of errors. It is only necessary to estimate the solvation Gibbs free energy of monomers and pairs. There are many methods, which offer calculation of solvation energies or Gibbs free energy. However, since different treatments often lead to contradictory results, the validation against available experimental data seems to be indispensable. Since there are not precise data characterizing pure stacking[30] experimental data for nucleobase solvation and of some heterocyclic compounds are used for a validation procedure. In Table S1 (ESI†) values of gas phase basicities and pK_a were collected. These data allow for estimation of solvation Gibbs free energies of monomers based on the above presented thermodynamic cycle. Among many available methods the standard Polarizable Continuum Model (PCM)[28,29] was used here along with Bondi's parameterization and explicit hydrogen atoms on B3LYP/aug-cc-pvDZ level. Although solvation Gibbs free energies obtained on such a way do not provide the direct thermodynamic characteristics of stacking in the condensed phase, mainly because of neglecting of entropic effects, they still serve as a qualitative measure of dehydration energetics.[31] Besides, it is not possible to compare experimental and theoretical predictions since the former do not separate different thermodynamic contributions to nucleobase pair formation.[30] The Gaussian software package[32] was applied for single point calculations for compounds belonging to the 'training set' as well as for all of analyzed pairs in stacked conformations. The accuracy of the applied PCM method is demonstrated in Fig. 9. As one can see, the agreement between experimental and estimated values of ΔG^{sol} is quite satisfactory, for the set of analyzed heterocyclic compounds.

Thus, this particular method seems to be suitable for description of solvation of nucleobases and it is reasonable to expect that also pairs in stacked conformations are to be described at least qualitatively correct. The estimated mean values of $\Delta\Delta G^{\mathrm{sol}}$ along with corresponding standard deviations are collected in Table 1. All values of relative solvation Gibbs free energy are positive, which agrees with chemical intuition that stacked conformations are thermodynamically

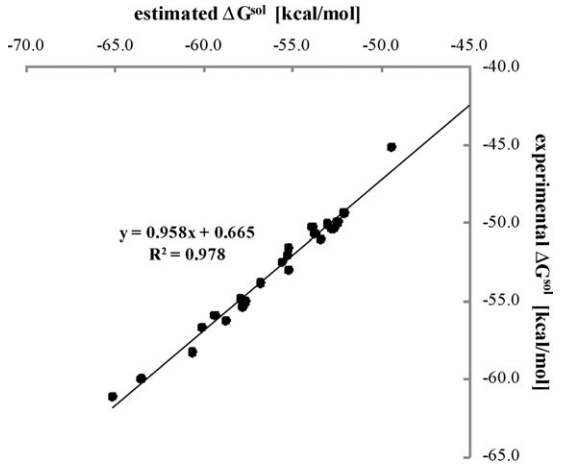


Fig. 9 The accuracy of solvation Gibbs free energies predicted for a training set comprising 27 heterocyclic compounds listed in Table S1 (ESI†). The PCM/Bondi method on B3LYP/aDZ level was applied.

less favorable in solution than solvated non-stacked monomers. Since guanine is the most polar nucleobase the desolvation associated with stacking of this particular base results in the highest values of $\Delta\Delta G^{\mathrm{sol}}$. There is observed correspondence between $\mathrm{IIE}^{\mathrm{MFO}}$ and $\Delta\Delta G^{\mathrm{sol}}$ although with only modest linear correlation ($R^2 = 0.69$). The stronger is the stabilization effect of stacking interactions in the gas phase the higher the values of relative solvation Gibbs free energy. Thus, the highest destabilization of stacking in water environment is expected for strongest stacked complexes such as G/A, A/G, G/C and C/G. On the contrary two stacked cytosine molecules are characterized by almost three-fold lower values of $\Delta\Delta G^{\mathrm{sol}}$. Such large discrepancies originate from the different nature of stacking interactions in these extreme cases. Interestingly the estimated mean values of $\Delta\Delta G^{\mathrm{sol}}$ are not context related since solvation of X/Y and Y/X complexes is almost identical (within values of standard deviation). These conclusions agree with those previously reported[31] despite the fact that the values of relative Gibbs free energies presented in Table 1 are higher than those obtained by Šponer *et al.*[31] One possible explanation, apart from the method used, is significant structure-related energetics also including solvation effects. Presented averaged values correspond to a much broader set of conformations. Additionally, contrary to prediction of Šponer *et al.*[31] C/C pairs are characterized by the smallest desolvation penalty among all studied sequences. Thus, good intrinsic stacking is not a direct indicator of the desolvation term. Also there are not correlations between polarities of stacked pairs and relative Gibbs free energies of solvation. Since, there is very good correlation between dipole moments estimated for gas phase and aqueous solution ($\Delta\Delta G^{\mathrm{sol(water)}} = 1.618\Delta\Delta G^{\mathrm{sol(gas)}} - 0.035$, $R^2 = 0.997$) the values of dipole moment related to either polar or non-polar environments do not directly inform about de-hydration penalties of stacked pairs.

Conclusions

Stacking as one of the fundamental intermolecular interactions in DNA double strands has attracted great attention in the last decade. However, there is no unequivocal

quantification of these interactions since the local conformations related to DNA flexibility affect significantly energetics of nucleobase pairs. Unfortunately, the relationships between structural parameters defining mutual orientation of two bases and corresponding interaction energies are not known due to the very complicated and highly non-linear nature. In this paper the methodology was applied for assessment of order of base–base stacking interactions in conformations matching to ones occurring in B-DNA crystals. The most frequently occurring energies of stacked pairs seem to have much more meaning than any particular conformation selected among many occurring in the polynucleotide double-stranded helices. Obtained distributions allow also for estimation of standard deviations and the range of interactions occurring in B-DNA crystals. Although such an extensive approach is rather time consuming, since several single point energy estimations must be performed, in the author's opinion it provides more general information about the intermolecular interactions in B-DNA than previously reported. Although it is possible to identify the most representative structures based on clustering protocols, due to observed significant structural heterogeneities of stacked complexes, the IIE assessment for the whole population seems to be beneficial. Of course it is important to bear in mind that energy itself is not the only component to the thermodynamic stabilization. Unfortunately, the inclusion of thermal corrections and entropy for obtaining values of Gibbs free energy is not a trivial task.[33] It is well known[34] that the MP2 method poorly predicts thermodynamic properties. Also, the accuracy of molecular mechanics methods, as for example the Poisson–Boltzmann surface area (MM-PBSA) continuum-solvent technique, also suffer from serious uncertainties. Thus, methods for reliable free-energy estimation are still to be developed. Fortunately the energetics of stacking interactions provide valuable insight into the intrinsic contributions to DNA stabilities, and collection of sufficient amount of data on the configurational hyperspace seems to be useful for formulation of the general model of structure–energy relationships in terms of non-linear theories, as for example for neural networks.

Acknowledgements

The results were partly obtained based on the computational grants from PCSS (Poznań Supercomputing and Networking Centre, Poland). The allocations of the CPU time is greatly appreciated.

References

1 H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan and B. Schneider, *Biophys. J.*, 1992, **63**, 751.
2 P. Hobza and J. Šponer, *Chem. Rev.*, 1999, **199**, 3247.
3 K. Muller-Dethlefs and P. Hobza, *Chem. Rev.*, 2000, **100**, 143.
4 P. Jurečka, J. Šponer, P. Černy and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985.
5 K. E. Riley and P. Hobza, *J. Phys. Chem. A*, 2007, **111**, 8257.
6 J. Šponer, K. E. Riley and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2595.
7 P. Jurečka, P. Nachtigall and P. Hobza, *Phys. Chem. Chem. Phys.*, 2001, **3**, 4578.
8 P. Cysewski, *J. Mol. Struct. (THEOCHEM)*, 2008, **865**, 36.
9 P. Cysewski, *J. Mol. Mod.*, 2009, **15**, 597.
10 P. Cysewski and P. Czeleń, *J. Mol. Model.*, 2009, **15**, 607.
11 P. Jurečka and P. Hobza, *J. Am. Chem. Soc.*, 2003, **125**, 15608.
12 P. Cysewski and Ż. Czyżnikowska-Balcerak, *J. Mol. Struct. (THEOCHEM)*, 2005, **757**, 29.
13 P. Cysewski and Ż. Czyżnikowska, *J. Heterocycl. Chem.*, 2007, **44**, 765.
14 P. Cysewski, Ż. Czyżnikowska, R. Zaleśny and P. Czeleń, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2665; S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.
15 H. J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, P. Celani, T. Korona, G. Rauhut, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, A. W. Lloyd, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, P. Palmieri, R. Pitzer, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni and T. Thorsteinsson, MOLPRO, Revision 2006.0, Patch(2006.1), Cardiff, UK, 2006.
16 H. B. Mann and D. R. Whitney, *Ann. Math. Statistics*, 1947, **18**, 50.
17 P. Jurečka, J. Šponer and P. Hobza, *J. Phys. Chem. B*, 2005, **108**, 5466.
18 Ż. Czyżnikowska, R. Zaleśny, M. Ziółkowski, R. W. Gora and P. Cysewski, *Chem. Phys. Lett.*, 2007, **450**, 132.
19 S. Neidle, *Nucleic Acid Structure and Recognition*, Oxford University Press, Oxford, 2002.
20 D. Svozil, J. Kalina, M. Omelka and B. Schneider, *Nucleic Acids Res.*, 2008, **36**, 3690.
21 M. J. Packer, M. P. Dauncey and C. A. Hunter, *J. Mol. Biol.*, 2000, **295**, 71.
22 C. A. Hunter and X.-J. Lu, *J. Mol. Biol.*, 1997, **265**, 603.
23 H. Cybulski and J. Sadlej, *J. Chem. Theory Comput.*, 2008, **4**, 892.
24 G. Hill, G. Forde, N. Hill, W. A. Lester, W. A. Sokalski and J. Leszczyńki, *Chem. Phys. Lett.*, 2003, **381**, 729.
25 J. G. Hill, J. A. Platts and H.-J. Werner, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4072.
26 R. R. Toczyłowski and S. M. Cybulski, *J. Phys. Chem. A*, 2003, **107**, 418.
27 P. Cysewski, *J. Mol. Model.*, 2007, **13**, 801.
28 B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **106**, 5151.
29 M. Cossi, G. Scalmani, N. Rega and V. Barone, *J. Chem. Phys.*, 2002, **117**, 43.
30 J. SantaLucia and D. Hicks, *Annu. Rev. Biophys. Chem.*, 2004, **33**, 415.
31 J. Šponer, P. Jurečka, I. Marchan, F. J. Luque, M. Orozco and P. Hobza, *Chem.–Eur. J.*, 2006, **12**, 2854.
32 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 and J. A. Pople, *GAUSSIAN 03 (Revision A. 1)*, Gaussian, Inc., Wallingford, CT, 2004.
33 J. Florian, J. Sponer and A. Warshel, *J. Phys. Chem. B*, 1999, **103**, 884.
34 A. P. Scott and L. Radom, *J. Phys. Chem.*, 1996, **100**, 1650.