J Mol Model (2009) 15:597–606
DOI 10.1007/s00894-008-0378-9

ORIGINAL PAPER

The post-SCF quantum chemistry 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

Piotr Cysewski

Received: 15 August 2008 / Accepted: 31 October 2008 / Published online: 28 November 2008

Abstract The energies of intra- and inter-strand stacking interactions in model d(GpC) and d(CpG) two-base-pair steps were estimated by MP2/aug-cc-pVDZ single point calculations corrected for basis superposition errors. The stacked two-nucleobase pairs were constructed using experimental values of base pair and base step parameters taken from Nucleic Acid Database (http://ndbserver.rutgers.edu/). Three distinct polymorphic forms were analysed, namely A-, B- and Z-DNA. The applied methodology enables statistical analysis of structural and energetic diversities. The structural relationships between polymorphic forms are quite complex and depend on the sequence of pairs. The variability of parameters such as *shift* and *tilt* is almost the same irrespective of the polymorphic form and sequence of steps analysed. In contrast, *shift* and *twist* distributions easily discriminate all three polymorphic forms of DNA. Interestingly, despite significant structural diversities, the energies of the most frequent energy ranges are comparable irrespective of the polymorphic form and base sequence. There was observed compensation of inter- and intra-strand interactions, especially for d(GpC) and d(CpG) steps found in A- and B-DNA. Thus, among many other roles, these pairs act as a kind of energetic buffer, balancing the double helix.

Keywords Cytosine · DNA · Guanine · Intermolecular interactions · Polymorphism · Stacking

Introduction

The DNA double helix is stabilised by a variety of different interactions [1–3]. The electrostatic and dispersion contributions play a distinct role in preserving the functionality of this polynucleotide in the cell environment [3]. Electrostatic interactions are dominant not only in the case of typical columbic interactions between charged groups, but also in all those contacts that involve permanent dipoles. For example, DNA molecules are stable in solution only at appropriate salt concentrations, since dissociating cations counterbalance the negative charge of the phosphate groups and prevent spontaneous denaturation in low ionic strength conditions [2]. Another source of the electrostatic contribution is hydrogen bonding of nucleobases. The permanent dipole moment of these heterocyclic compounds keeps the opposite strands in equilibrium. In contrast, dispersion comes mainly from the stacking interactions that maintain the helical structure [4]. Thus, it is of crucial importance to quantify these types of contribution to the stabilisation energy of the DNA double helix. Although stacking of nucleobases has been the subject of numerous studies [3–15], many questions still remain unanswered. Due to the activities of the Hobza group [3–11] and others [12–15] significant insights into the nature and role of stacking interactions has been gained. In addition, an accurate methodology has been validated for many stacked complexes [3, 4]. It is commonly accepted that advanced quantum chemistry post-SCF methods are the only appropriate tools for quantitative description of stacking interactions.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-008-0378-9) contains supplementary material, which is available to authorized users.

P. Cysewski (✉)
Department of Physical Chemistry, Collegium Medicum,
Nicolaus Copernicus University,
Kurpińskiego 5,
85-950 Bydgoszcz, Poland
e-mail: piotr.cysewski@cm.umk.pl

Although the CCSD(T) level is recommended [4], it is extremely demanding and impractical for routine applications. On the other hand, modest electron correlation treatments like MP2 still provide valuable information since they cover a significant portion of electron correlations. This is due to "fortunate errors compensation" since higher order contributions to electron correlation that are not included in the MP2 approximation are typically positive, while extension of the basis set increases the dispersion attraction [4, 16–18]. Thus, it is reasonable to assume that MP2/aDZ values of base-stacking energies are not very far from the actual values.

Stacking interactions critically depend on nucleobase orientation [3, 12]. This is particularly important from the perspective of DNA double strand polymorphism. Sequence dependence is another factor affecting the energy of stacking complexes. Thus, selection of conformations for quantum chemistry studies is crucial, and is no trivial task. Many trials applying different scanning protocols [12–14], although providing valuable data, do not deal with the polymorphism and sequence-dependence of stacking interactions. Besides, sampling of the configuration hyperspace of two stacked nucleobases pairs is a tremendous task and any brute force method is inappropriate. Since the aim of this paper is to provide direct insight into the distribution of stacking interactions in model d(GpC) and d(CpG) steps in conformations corresponding to different forms of DNA helices, the structural information used is taken directly from X-ray diffraction patterns of DNA crystals. The Nucleic Acid Database (NDB; http://ndbserver.rutgers.edu/) [19] provide comprehensive details on atomic resolution. The typical set consists of base pair (*shear*, *stretch*, *stagger*, *buckle*, *propeller* and *opening*) and base step (*shift*, *slide*, *rise*, *tilt*, *roll* and *twist*) parameters. Nowadays, this procedure has been successfully applied to G/G stacking [20]. Here, the interaction of guanine with cytosine is analysed in detail to provide a description of the polymorphism-related and sequence-dependent heterogeneity of base-base stacking interactions in A-DNA, B-DNA and Z-DNA.

Methods

X-ray diffraction images of DNA crystals deposited with NDB [16] were the source of all 18 structural parameters defining the intrinsic structure of d(CpG) and d(GpC) base pair steps. The mutual orientation of two base pairs in double helical DNA is univocally defined in the standard reference frame [21] by providing two sets of base pair parameters and one set of base step parameters. The base pairs were prepared as described in previously [20] and only a brief summary is provided here. Among all available DNA structures, only those related to native double helices without any ligands/complexes with metal ions or proteins, with no mismatches or any chemical modifications of bases, sugar or phosphate moieties were taken into account. A list of all structures analysed is provided in supporting materials (Table S1). The program X3DNA [19] was used for preparation of base pairs of given sequence and orientation based on all 18 parameters provided by the PDB files. All atoms were removed except those belonging to stacked pairs. In such a way, the sugar-phosphate backbones were simplified to just hydrogen atoms. These input files were used for single point energy calculations at DF-MP2/aug-cc-pvdz (aDZ) level of theory. The main advantage of this procedure is that prepared pairs exactly match structures found in DNA crystals, but monomers have geometries optimised at the MP2/aDZ level in Cs symmetry. The counterpoise correction for BSSE error [22] was included in all single point calculations of stacked complexes. The MolPro package [23] was used in quantum chemistry calculations. The above procedure leads to model d(CpG) and d(GpC) steps. Their energetics are characterised assuming a pairwise additive character of all base–base interactions. This is justified by the commonly accepted assumption that many-body contributions are rather small compared to two-body terms [18, 24].

Results and discussion

Two aspects of the inter- and intra-strand stacking interactions in d(CpG) and d(GpC) steps found in B-DNA, A-DNA and Z-DNA crystals were analysed in detail. Firstly, some comments on structural heterogeneity are provided in relation to d(GpC) and d(CpG) pairs found in the three analysed polymorphic forms of DNA. Secondly, the intermolecular interaction energies (IIE) occurring in these steps are characterised. The most important feature of this study is the statistically meaningful amount of data used in the analysis. Thus, instead of providing energy values related to particular pairs of stacked nucleobases, a distribution analysis is presented. The structural heterogeneities and related variability of IIEs are presented as smoothed histograms. The plots were drawn by dividing the whole range of a particular variable into ten equal intervals, and calculating the percentage of the population within each period. Although there are ten unique sequences of stacked nucleobase pairs in DNA, not all of them are encountered in biologically important polymorphic forms of nucleic acid. For example, selected Z-DNA structures contain the highest percentage (56%) of d(CpG) steps, and d(GpC) sequences constitute about 41% of all available pairs in Z-DNA; the remaining stacked pairs are present in only marginal amounts in this polymorphic form of DNA. On the contrary, the concentration of d(GpC) and d(CpG) sequences in selected B-DNA structures is much smaller, at

about 12% and 19%, respectively. A-DNA comprises about 14% of both base pairs of guanine and cytosine. Thus, only guanine and cytosine stacking is available in all three polymorphic forms analysed, thus justifying the subject of this study.

Before presenting and discussing IIE values, some important comments regarding the interpretation of the data are required. First of all, gas phase calculations do not directly reflect the energetics of nucleic acids, and extrapolation to bulk phase is not straightforward. This is a complex issue that has been discussed extensively in the literature [4, 25]. X-ray crystal diffraction patterns and NMR experiments reveal structures at atomic resolution but they do not provide direct insights into energetics. Although free energy values can be obtained by thermodynamic experiments in condensed-phase, they do not directly provide geometries. Quantum chemistry studies may then offer a very valuable bridge between these two types of experiments. Unfortunately, the direct correlation between gas-phase and condensed-phase data is often weak [25], and there is no simple method for deducing experimental stabilities from gas-phase data, or, conversely, predicting the intrinsic forces from the condensed-phase data. Despite all these limitations, gas-phase data may still be helpful in proper interpretation of many aspects of the experimental data. For example, there is a complex interplay between the intrinsic base–base terms and the resulting nucleic acid's structure and stability. The molecular forces in nucleic acids are so variable, and so related to structural context, that a certain type of interaction may have a completely different effect on DNA stability in different situations. The d(GpC) and d(CpG) steps analysed below reveal such intricate interrelations. Without the details that can be exposed via advanced quantum chemistry computations, our understanding of the energetics of nucleic acids would be limited even if precise condensed-phase data are available. It must also be emphasised that post-SCF methods, including correlation effects, basis superposition errors and extended basis sets, usually provide energies only at zero Kelvin temperatures. The available condensed phase data [26–28] offer free energies at various temperatures. Attempts [29–31] to include the influence of the environment on stacking have encountered many methodological and technical problems that have been only partly solved until now. Thus, only raw intermolecular interaction energies are presented and discussed here.

Structural diversities of polymorphic d(GpC) and d(CpG) steps

It is commonly known that DNA can exist in many conformations [1, 2, 32–35] but only forms such as A-DNA, B-DNA, and Z-DNA have been observed in living organisms. The conformation of DNA depends on many factors, such as nucleotide sequence, solution conditions, concentration of metal ions, the amount and direction of supercoiling, chemical modification of the bases, and many others [36]. The B-form—a right-handed spiral with two grooves of different size—is the most common under the conditions found in cells [1, 2, 36]. A-DNA is also a right-handed double helix and is fairly similar to B-DNA. However, the A-form has a shorter and more compact helical structure. The increase in the number of base pairs per turn results in deepening of the major groove, making the minor groove shallower [1, 2]. Z-DNA is very different from other DNA forms, being a left-handed double helical arrangement in which the double helix twists in a zig-zag pattern [1, 37] via the alternation of two different dinucleotide conformations with either a large *twist* and a small *slide* or a small *twist* and a large *slide* between adjacent base pairs. This results not only in poor stacking within the latter dinucleotide repeat but also allows for the direct contacts of the O4′ atom of 2′-deoxyribose with the pyrimidine ring of guanine [18, 38]. Although the stabilising effect of such an interaction is comparable to normal base–base stacking, this contribution is not studied here.

The structural characteristics of different polymorphic forms of DNA can be obtained directly from the NDB database [19]. The values of base pair and base step parameters may be taken directly from the website (http://ndbserver.rutgers.edu/) or recalculated using the X3DNA program based on available PDB files. The latter method was used in this paper. Since the aim of this study was to characterise stacking interactions, only the distributions of base step parameters are provided in Figs. 1 and 2. These smoothed histograms demonstrate the most important sequence-dependent and polymorphic from-related diversities of displacement (*shift*, *slide*, *rise*) and angular (*tilt*, *roll*, *twist*) parameters corresponding to the crystals structures of B-, A- and Z-DNA. For example, the distribution of *shift* values are very similar for both d(GpC) and d(CpG) steps in all three DNA forms. Although all distributions are statistically different ($P<0.001$), their range, shape and localisation of maximums are comparable. The *shift* values are the measure of mutual displacement of two successive base pairs along the shorter molecular axis. Striking similarities in nucleobase shifting is observed for the steps analysed irrespective of the polymorphic form. The distributions of *slide* values presented in Fig. 1 reveal the very different structures of all three polymorphic forms of DNA. This parameter quantifies mutual displacement of two successive base pairs along the line formed by the longer molecular axis. In the case of the d(CpG) sequence, Z-DNA adopts positive and much higher values of this parameter than A- or B-DNA. Most d(CpG) pairs found in the A-form are characterised by the smallest and most

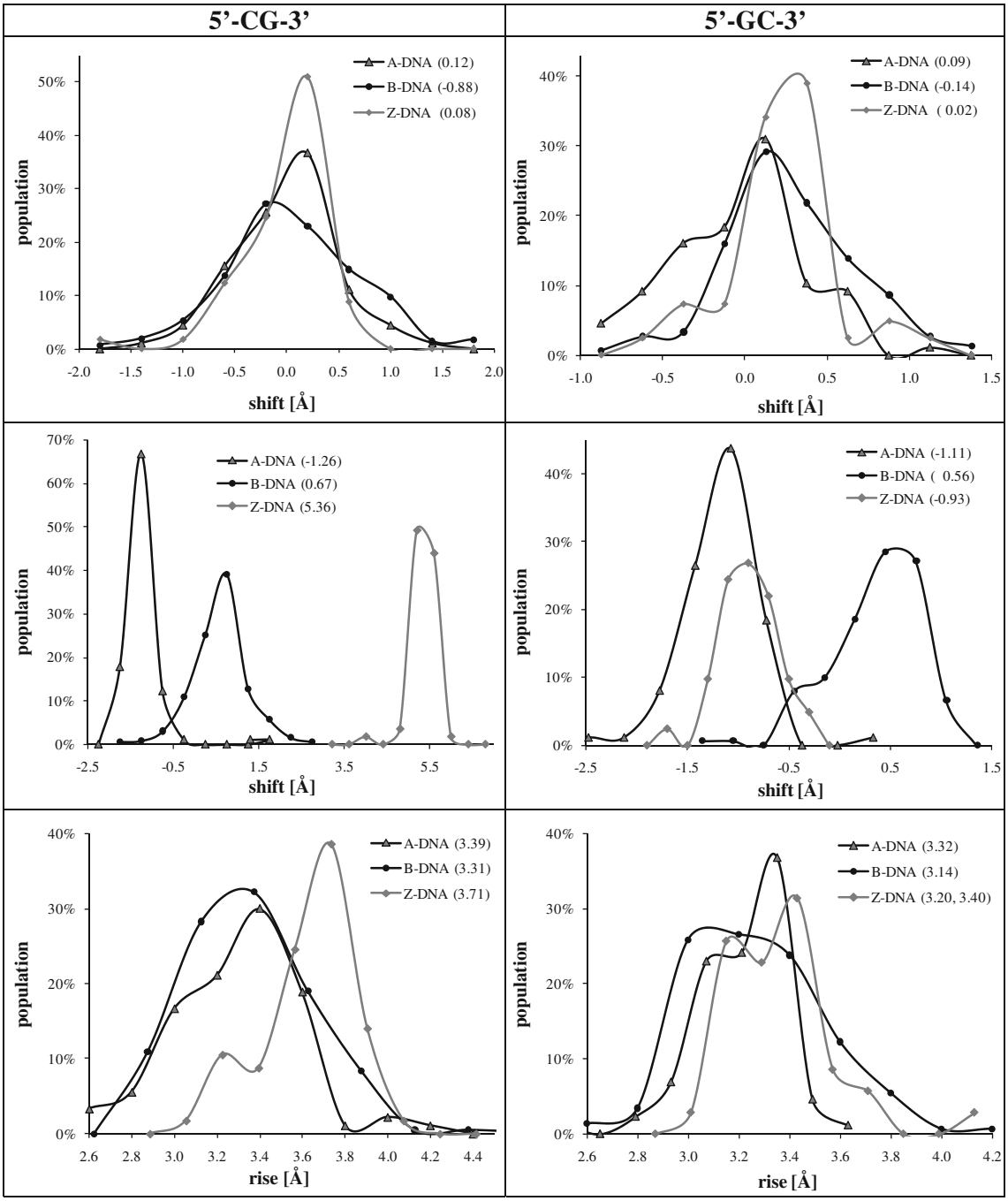


Fig. 1 Smoothed histograms presenting the distributions of displacement parameters (*shift*, *slide*, *rise*) corresponding to GC and CG doublets found in different polymorphic forms of crystallised DNA [19]. Values in parenthesis define localisation of maximums on the histogram plots

negative values of *slide* among all the steps analysed in this paper. In the case of d(GpC) sequences, the differences between Z- and A-DNA are not so great, and distributions of *slide* for pairs in the Z-form are more similar to those of A-DNA than those of B-DNA.

The third parameter analysed, *rise*, defines mutual displacement of two successive base pairs along the axis perpendicular to the molecular plane. As may be inferred from Fig. 1, the separation between stacked bases is highest for d(CpG) steps in Z-DNA. Both d(CpG) and d(GpC) sequences have almost identical *rise* distributions for A-DNA and B-DNA. The fourth parameter analysed, *tilt*, belongs to the set of angular variables and is defined by the rotation of two successive base pairs about the shorter molecular axis. As shown in Fig. 2, all polymorphic forms are typified by very similar distributions of that parameter,

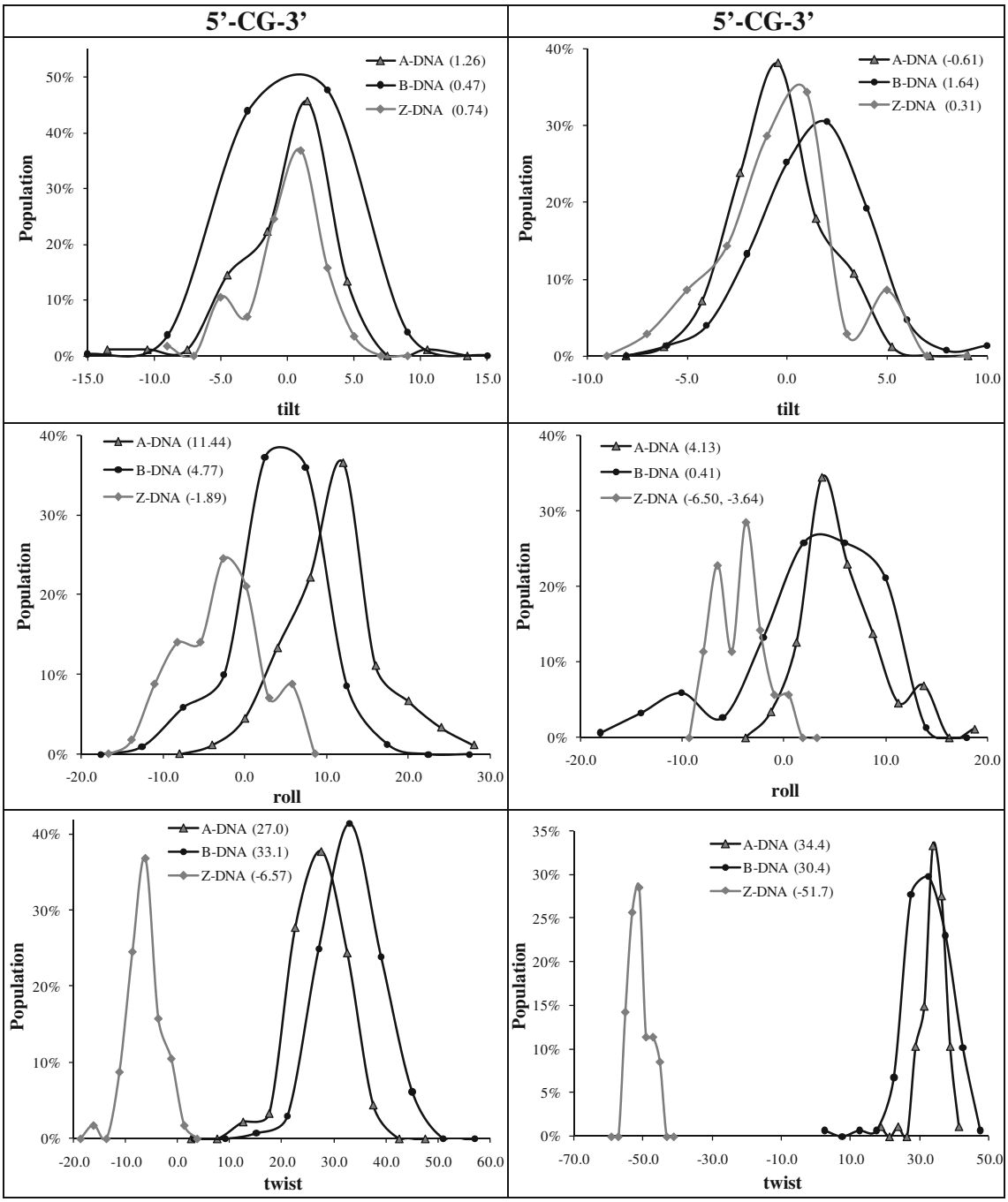


Fig. 2 Smoothed histograms presenting the distributions of angular parameters (*tilt*, *roll*, *twist* expressed in degrees) corresponding to GC and CG doublets found in different polymorphic forms of DNA [19]. Values in parenthesis define localisation of maximums on the histogram plots

irrespective of the sequence of guanines and cytosines. This is not the case for *roll* values, since the rotation of two successive base pairs about the longer molecular axis is very similar for A- and B-DNA, but is significantly smaller for Z-DNA. This is valid both for d(CpG) and d(GpC) pairs. The lowest, and negative, values of *roll* found in Z-DNA indicate the significant rotation of bases belonging to strand II. In contrast, higher rolling of bases from strand I is observed for both A- and B-DNA. In d(CpG) steps, this effect is more pronounced for B-DNA.

The last parameter analysed, *twist*, describes the rotation of two successive base pairs about the axis perpendicular to the plane formed by the hydrogen-bonded pair. As is clearly evident in Fig. 2, the Z-DNA structure differs significantly from other DNA forms. Negative values of *twist* characterise both d(CpG) and d(GpC) pairs in this DNA form. However,

there are significant sequence dependencies of the *twist* distributions. The most frequently adopted values of *twist* in the Z-form are about −6.6° and −51.7° for d(CpG) and d(GpC) pairs, respectively. This discrepancy in twisting is responsible for the different overlapping of the stacking bases, which has consequences for the intermolecular interactions. Differences in *twist* distributions for A- and B-DNA are also observed, but the divergence is not as great. In the case of d(CpG) pairs, the maximums of the histogram plots are found for slightly smaller values of *twist* for A-DNA compared to the B-form. This order is reversed for d(GpC) steps.

This short analysis of structural parameters demonstrates that discrimination of the polymorphic forms is significantly sequence dependent. However, not all base step parameters are characterised by distinct distributions for all DNA double helices analysed. The variability of such parameters as *shift* and *tilt* is almost the same irrespective of the polymorphic form and sequence of the analysed pairs. On the contrary, *shift* and *twist* distributions easily discriminate all three polymorphic forms of DNA. The sequence dependence is rather modest in the case of A- and B-DNA, except for *slide* values. Thus, the structural relationships between polymorphic forms are quite complex and depend on the sequence of the pairs.

Energetic heterogeneities of intermolecular interactions

Different kinds of intermolecular interactions exist between heterocyclic nucleobases in double-stranded DNA. Apart from hydrogen bonding, which is not discussed here, there is intra- and inter-strand stacking. All these interactions may be sequence-dependent, as presented schematically in Fig. 3. Obviously, two distinct pairs define the intra-strand stacking of guanine with cytosine, namely 5′-G/C-3′ and 5′-C/G-3′. Due to the symmetry of the d(GpC) and d(CpG) steps, they may occur both in strand I and strand II. However, irrespective of the polymorphic form of DNA, these kinds of interactions are structurally and energetically equivalent and only two types of intra-strand stacking between guanine and cytosine are to be considered. The inter-strand interactions may be formed in four distinct ways, as presented in Fig. 3. All of these stacking interactions may occur in B-DNA, A-DNA and Z-DNA and must be analysed separately. Thus, the full characteristics of all stacking interactions calls for analysis of 12 sets of pair arrangements, which takes into account both the polymorphism and sequence-dependence of nucleotide pairs.

Intra-strand stacking between guanine and cytosine

Figure 4 presents the distributions of the intra-strand interactions for all three analysed polymorphic forms of DNA. Several interesting features may be inferred from these data. First of all, a broad spectrum of stacking interactions may be observed in crystallographic DNA. There are very strong interactions exceeding −12 kcal/mol, but on the other hand there are many structures for which the intra-strand stacking is very weak. Although these extreme values are relatively rare, they still contribute to the total diversity of IIE in DNA of different polymorphic forms. In addition, significant sequence-related differences in the energies of intermolecular interactions are observed. Taking into account the most frequently occurring values of intermolecular interactions, one may conclude that, irrespective of the polymorphic form, stacking in G/C pairs is energetically more favourable than stacking in C/G pairs. In fact, on the whole, IIE distributions corresponding to G/C pairs are significantly shifted toward stronger interactions in comparison to C/G pairs (Fig. 4). In addition, the polymorphism of DNA double strands has a considerable

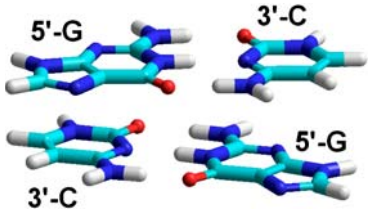
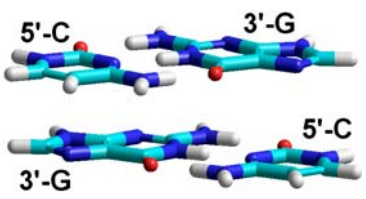
model d(GpC) step	model d(CpG) step
	
intra-strand interactions: 5'-G/C-3' in (I) and (II) strands inter-strand interactions: 5'-G|G-5' and 3'-C|C-3'	intra-strand interactions: 5'-C/G-3' in (I) and (II) strands inter-strand interactions: 5'-C|C-5' and 3'-G|G-3'

Fig. 3 Classification of potential intermolecular interactions that may occur in stacked two-base-pairs formed by guanine and cytosine in DNA double stranded helices. *Diagonal* and *vertical lines* denote intra- and inter-strand stacking interactions, respectively (e.g. 5′-G/C-3′; 3′-C|C-3′)

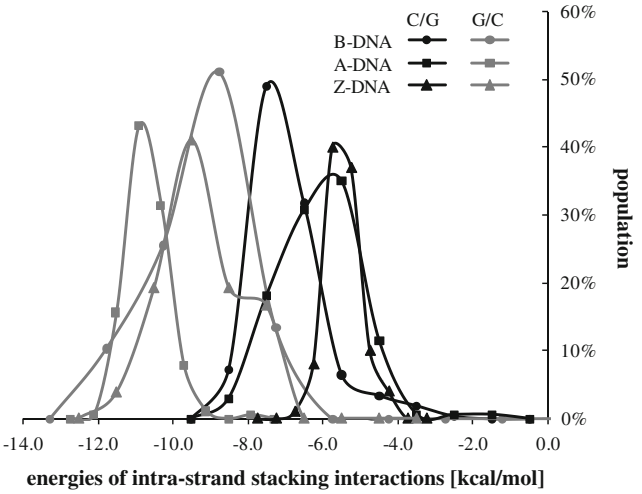


Fig. 4 Intra-strand stacking interactions of guanine with cytosine in different polymorphic forms of crystallographic DNA double helices

impact on inter-strand stacking interactions. Unfortunately, no simple interpretation of that factor is possible since it influences G/C and C/G pairs in different ways. For example, according to the position of the maximums on the smoothed histograms, C/G interactions in B-DNA are stronger than in A-DNA. On the contrary, the energy of stacking in G/C pairs is reversed for B- and A-DNA. Values corresponding to medians of IIE distributions are presented in Table 1. It is noteworthy that all distributions are statistically distinct ($P<0.001$). However, in some cases the median values corresponding to different polymorphic forms of the same sequence differ by less than the standard deviation. Taking into account the distributions of intermolecular interaction energies, one may conclude the following sequence for intra-strand stacking between guanine and cytosine: G/C(A-DNA) > G/C (Z-DNA) ≈ G/C(B-DNA) > C/G (B-DNA) >C/G (A-DNA) > C/G (Z-DNA).

Inter-strand stacking in cytosine–cytosine pairs

Two cytosine molecules belonging to opposite strands may interact via two different motifs depending on the pair sequence (see Fig. 3). As shown in Fig. 5, the 5′-C|C-3′ and 3′-C|C-3′ interactions are energetically distinct for any polymorphic form of DNA. Apart from sequence-dependence, DNA conformation significantly affects the IIE distributions. Interestingly, unlike with intra-strand stacking, most cytosine–cytosine inter-strand interactions are repulsive in character. The only exception observed is for 5′-C|C-5′ pairs found in Z-DNA. In this case, the attractive contribution is characterised by a median value equal to −3.0 kcal/mol. As listed in Table 1, the other IIE distributions of C|C pairs are characterised by positive median values. Furthermore, these values do not differ significantly from each other since all are within the corresponding standard deviation. There is a rather simple structural rationale for the observed nature of C|C interactions. In the case of 3′-C|C-3′, the monomers are rather far each from other and the distance between heavy atoms usually exceeds 4 Å. This is common for all three analysed DNA forms. A similar property is also typical of 5′-C|C-5′ pairs in A- and B-DNA. On the contrary, the spatial distribution of 5′-C|C-5′ pairs in Z-DNA is completely different since overlapping of the side groups of one cytosine molecule with the heterocyclic ring of another occurs. Thus, a significant electrostatic contribution to the interaction energies is expected as a result of direct π–n interactions, which is the source of the observed attraction in this case. The different orientations of cytosine molecules are related mainly to *slide* values. As mentioned above, the d(CpG) steps in Z-DNA are characterised by positive and by very high values of this parameter. This makes direct 5′-C|C-5′ interactions possible but only if DNA adopts the Z-form. It is noteworthy that the repulsive contribution of most C|C pairs is much smaller than most

Table 1 The statistical characteristics of inter- and intra-strand stacking interactions occurring in model d(CpG) and d(GpC) steps in different polymorphic forms. All median and standard deviation values of intermolecular interactions energies are expressed in kcal/mol. Total d(CpG) and total d(GpC) denote the sum of the inter-strand, intra-strand stacking contributions and hydrogen bonding

	B-DNA Median	A-DNA	Z-DNA	B-DNA Standard deviation	A-DNA	Z-DNA
5′-C/G-3′	−7.1	−6.1	−5.5	1.0	1.1	1.3
5′-G/C-3′	−9.0	−10.7	−9.6	1.2	0.5	1.0
5′-C|C-5′	1.0	1.1	−3.0	1.0	0.2	0.6
3′-C|C-3′	2.6	2.5	1.5	0.3	0.5	0.5
5′-G|G-5′	−0.4	2.3	2.3	1.1	0.5	0.4
3′-G|G-3′	−3.7	−5.7	0.8	0.8	1.2	0.2
total CG	−65.8	−60.0	−59.0	5.83	4.12	3.57
total GC	−67.8	−58.9	−61.0	7.33	6.49	4.50

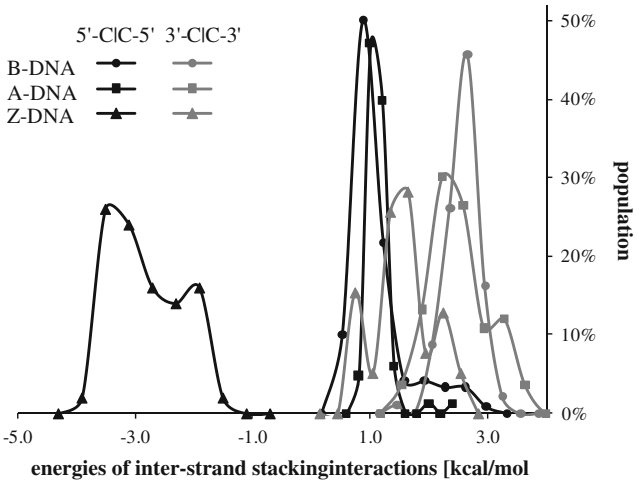


Fig. 5 Inter-strand cytosine–cytosine intermolecular interactions present in d(GpC) and d(CpG) steps in conformations corresponding to three different polymorphic forms of the DNA double helix

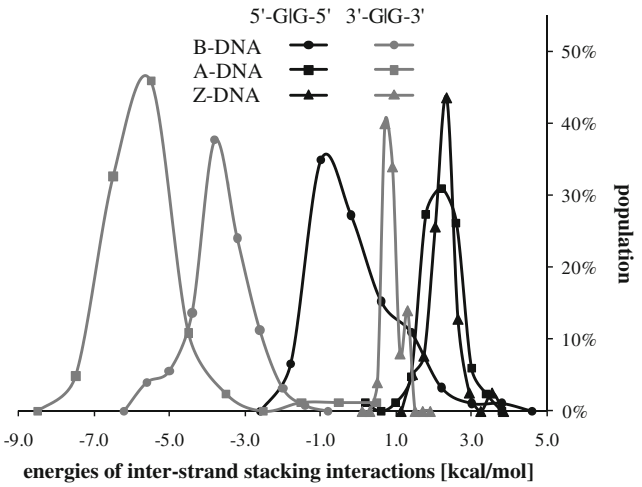


Fig. 6 Inter-strand guanine–guanine intermolecular interactions present in d(GpC) and d(CpG) steps in conformations corresponding to three different polymorphic forms of the DNA double helix

intra-strand stacking interactions in d(CpG) and d(GpC) steps. Thus, compensation of interactions occurs, thus stabilising all analysed two-base-pairs in any polymorphic form of DNA. In addition, there is very narrow window of C|C interactions and in most cases the structural heterogeneity affects the inter-strand interactions of two cytosine molecules only very slightly. The only exceptions are the 5′-C|C-5′ interactions found in Z-DNA, which differ significantly from the other analysed pairs. Most of these pairs are characterised by IIE values within −3.0±0.6 kcal/mol. The much broader range of 5′-C|C-5′ interactions suggests that the high elasticity and flexibility of the Z-form may be gained also due to these kind of interactions, since almost half of the pairs present in Z-DNA are CG sequences.

Inter-strand stacking in guanine–guanine pairs

In d(CpG) and d(GpC) steps, the two guanine molecules may also interact via inter-strand intermolecular interactions. Both DNA polymorphism and sequence alteration affect guanine–guanine inter-strand contacts. As illustrated in Fig. 6, G|G interactions are repulsive for Z-DNA irrespective of the sequence. The 3′-G|G-3′ pairs formed in A-DNA interact via the strongest forces among all inter-strand interactions, and they are as large as typical intra-strand stacking. Also, 3′-G|G-3′ interactions in B-form DNA are attractive for a very broad range of conformations. Other G|G interactions are repulsive in character. In particular, all d(GpC) pairs in which 5′-G|G-5 is present are characterised by considerable repulsion between two guanine molecules. In A-DNA, very strong G|G attractions result from fortuitous arrangements of monomers in d(CpG) pairs. Of all the polymorphic forms of nucleic acids studied here, A-DNA is characterised by the lowest values of *slide* and the highest values of *roll*. This allows for significant overlapping of the pyrimidine rings of two guanine molecules belonging to opposite strands. Thus, the π–π interactions typical of intra-strand stacking may also be manifested in this case. The same feature is also observed for d(CpG) pairs in B-DNA although the effect is much less pronounced in this case.

Additive character of stacking interaction energies

Inter- and intra-strand interactions are mutually related since, obviously, the more the monomers overlap, the stronger the interactions. Such partial covering of nucleobases may occur within the same strand, which consequently prevents overlapping of bases from the opposite strand. The reverse statement is also true, i.e. the more extended the positions of monomers in one strand, the stronger the inter-strand overlapping. This has direct consequences for the intermolecular interactions. Table 1 and Fig. 7 present a summation of values of all kinds of interactions between nucleobases in d(GpC) and d(CpG) steps; hydrogen bonding energies are also included. An interesting homogeneity is observed for pairs in B- and A-DNA conformations. For these two polymorphic forms, the most frequently occurring intermolecular interactions are almost sequence independent. Thus, the gain of attraction coming from intra-strand stacking is accomplished with the loss of IIE originating from inter-stand interactions. Thus, the repulsion interactions occurring in C|C and G|G are compensated by the attraction character of C/G and G/C stacking. This delicate balance between magnitude of attraction and repulsion makes stacked cytosine–guanine pairs almost iso-energetic in A-DNA, B-DNA and Z-DNA. The only exception to this rule is the stabilisation of d(GpC) steps in Z-DNA; of all the sequences studied in this

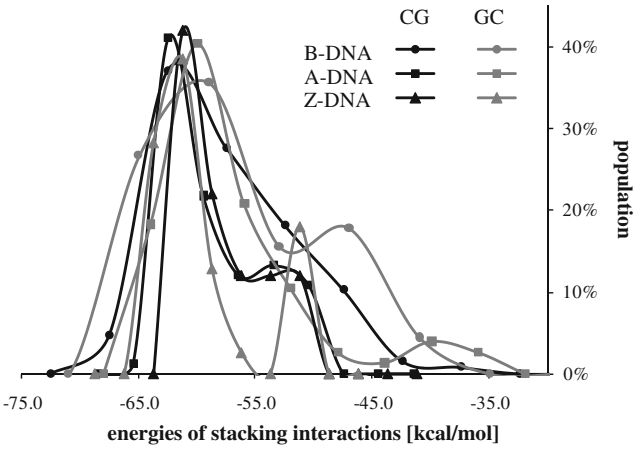


Fig. 7 Distributions of the sums of all intra-, inter-strand and H-bonding intermolecular interactions occurring in d(GpC) and d(CpG) steps present in different polymorphic forms of crystallographic DNA double helices

paper, these seem to be the least stable. As illustrated in Fig. 7, the corresponding IIE distribution differs from the other studied doublets. The data presented in Fig. 7 and Table 1 assume the pair-wise additive nature of intermolecular interactions. It is well known that the three-body and four-body terms are rather small [24] and of similar magnitude for different sequences. Thus, the two-body contribution determines the two base stacked pairs energetic.

Conclusions

The methodology applied in this paper relies on post-SCF ab initio quantum chemistry calculations of intermolecular interaction energies occurring in nucleic acid double helices. Instead of using conformations coming from particular model of DNA, the structural parameters defining actual arrangements of bases in d(GpC) and d(CpG) steps were used. Since the source of these structural parameters are X-ray diffraction experiments of crystallographic oligonucleotides, the statistics of experimentally observed structural and energetic diversities are available. The variability of the base step parameters is quite complex in relation to both the sequence and polymorphism of the DNA structures analysed. To the best of the author's knowledge, the energetic consequences of these structural diversities have not been studied to a reliable advanced ab initio level to date. The application of quantum chemistry methodology is indispensable in the proper description of the energies of weakly interacting systems, as has been stated by many previous studies [3, 4]. Thus, the results presented here provide unique and comprehensive data on the intermolecular interactions between guanine and cytosine in DNA. The most important conclusion is that the energy compensation effect occurs between different contributions of inter- and intra-strand interactions. Despite significant heterogeneity of the distributions of the base step parameters, the most frequent IIE values are very similar, irrespective of the polymorphic forms and base sequence. The only exception was observed for d(CpG) pairs in Z-DNA, which are significantly less stable than all other pairs analysed here. However, it is almost certain that interactions between sugar and guanine not included in this analysis significantly increase the stability of this sequence. The observed compensation of inter- and intra-strand interactions, especially apparent for guanine–cytosine pairs found in A- and B-DNA, justifies the selection of these particular building blocks by evolution. The great stability and flexibility of DNA is vital for its functionality. According to the results presented here, even strong structural fluctuations do not affect significantly the energy of d(GpC) and d(CpG) steps in A- and B-DNA. Thus, among many other roles, these sequences act as a kind of energetic buffer, balancing the double helix. The magnitude of intra-strand stacking energy is due mainly to π–π aromatic interactions and London dispersion contributions. Inter-strand stacking is governed mostly by the electrostatic contributions of the side groups. The compensation provided by these two distinct terms is probably the reason that such heterocyclic compounds were selected by nature for storing genetic information. It would be interesting if such compensation were also possible for other sequences existing in DNA or in other potential building blocks extending the genetic alphabet; however, such aspects are beyond of the scope of this study.

Finally, the observed spread of energy values presented above is deserving of comment. The local geometry variations of two base pairs directly affecting the heterogeneity of the intermolecular interaction are determined by internal and external effects: the former coming from oligomer sequence and local structure, while the latter is caused by crystal packing. However, numerous base pair steps characterized by poor energies are observed. These could originate from X-ray diffraction interpretation errors and refinement inaccuracies. The nominal resolution of X-ray data is not a sufficient indicator since it does not include possible inaccuracies or even incompleteness of the refinement procedures. Although base pair and step parameters rather than Cartesian coordinates were used in this work, experimental errors will still affect stacking energies. The visible right-hand tails on the presented IIE distributions are more likely to be related to experimental errors in structural parameters. Thus, gas-phase intermolecular interaction energy calculations may then serve as additional validation tools for experimentally derived structures. A detailed analysis of this interesting aspect will be addressed in a forthcoming paper.

Acknowledgements The technical assistance of Ms. A. Cieślińska is greatly appreciated. The results were partly obtained based on computational grants from PCSS (Poznań Supercomputing and Networking Centre, Poland).

References

1. Saenger W (1988) Principles of nucleic acid structures. Springer, Berlin
2. Bloomfield VA, Crothers DM, Tinoco I (1999) Nucleic acids: structures, properties and functions. University Science Books, Sausalito, CA
3. Hobza P, Šponer J (1999) Chem Rev 99:3247–3276. doi:10.1021/cr9800255
4. Šponer J, Riley KE, Hobza P (2008) Phys Chem Chem Phys 10:2595–2610. doi:10.1039/b719370j
5. Müller-Dethlefs K, Hobza P (2000) Chem Rev 100:143–167. doi:10.1021/cr9900331
6. Iyer RR, Pluciennik A, Burdett V, Modrich PL (2006) Chem Rev 106:302–323. doi:10.1021/cr0404794
7. Šponer J, Florian J, Ng H, Šponer J, Špacková N (2000) Nucleic Acids Res 24:4893–4902
8. Hobza P, Šponer J, Polášek M (1995) J Am Chem Soc 117:792–798. doi:10.1021/ja00107a023
9. Dąbkowska I, Jurečka P, Hobza P (2005) J Chem Phys 122:204322. doi:10.1063/1.1906205
10. Dąbkowska I, Gonzalez HV, Jurečka P, Hobza P (2005) J Phys Chem A 109:1131–1136. doi:10.1021/jp046738a
11. Jurečka P, Hobza P (2003) J Am Chem Soc 125:15608–15613. doi:10.1021/ja036611j
12. Cysewski P, Czyżnikowska Ż (2007) J Heterocycl Chem 44:765–773
13. Cysewski P, Czyżnikowska Ż, Zaleśny R, Czeleń P (2008) Phys Chem Chem Phys 10:2665–2672. doi:10.1039/b718635e
14. Cysewski P, Czyżnikowska Ż (2005) J Mol Struct THEOCHEM 757:29–36. doi:10.1016/j.theochem.2005.06.014
15. Sharma R, McNamara JP, Raju RK, Vincent MA, Hillier IH, Morgado CA (2008) Phys Chem Chem Phys 10:2767–2774. doi:10.1039/b719764k
16. Hobza P, Selzle HL, Schlag EW (1994) J Am Chem Soc 116:3500–3506. doi:10.1021/ja00087a041
17. Šponer J, Hobza P (1997) Chem Phys Lett 267:263–270. doi:10.1016/S0009-2614(97)00118-8
18. Šponer J, Gabb HA, Leszczynski J, Hobza P (1997) Biophys J 73:76–87
19. Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B (1992) Nucleic Acids Biophys J 63:751–759
20. Cysewski P (2008) J Mol Struct THEOCHEM 865:36–43. doi:10.1016/j.theochem.2008.06.019
21. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung CS, Westhof E, Wolberger C, Berman HM (2001) J Mol Biol 313:229–237. doi:10.1006/jmbi.2001.4987
22. Boys SF, Bernardi F (1970) Mol Phys 19:553–566. doi:10.1080/00268977000101561
23. Werner HJ, Knowles PJ, Lindh R, Manby FR, Schütz M, Celani P, Korona T, Rauhut G, Amos RD, Bernhardsson A, Berning A, Cooper DL, Deegan MJO, Dobbyn AJ, Eckert F, Hampel C, Hetzer G, Lloyd AW, McNicholas SJ, Meyer W, Mura ME, Nicklaß A, Palmieri P, Pitzer R, Schumann U, Stoll H, Stone AJ, Tarroni R, Thorsteinsson T (2006) MOLPRO, Revision 2006.0, Patch(2006.1), Cardiff, UK
24. Hill JG, Platts JA (2008) Phys Chem Chem Phys 10:2785–2791. doi:10.1039/b718691f
25. Šponer J, Jurečka P Hobza P (2006) In: Šponer J, Lankaš F (eds) Computational studies of RNA and DNA. Springer, Dordrecht, pp 343–388
26. Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH (1986) Proc Natl Acad Sci USA 83:9373–9377. doi:10.1073/pnas.83.24.9373
27. Turner DH, Sugimoto N, Kierzek R, Dreikert SD (1987) J Am Chem Soc 109:3783–3785. doi:10.1021/ja00246a047
28. Santalucia J (1998) Proc Natl Acad Sci USA 95:1460–1465. doi:10.1073/pnas.95.4.1460
29. Šponer J, Jurečka P, Marchan I, Luque FJ, Orozco M, Hobza P (2001) Chem Eur J 12:2854–2865
30. Meneni S, Shell SM, Gao L, Jurečka P, Šponer J, Lee W, Zou Y, Chiarelli MP, Cho BP (2007) Biochemistry 46:11263–11278. doi:10.1021/bi700858s
31. Florian J, Šponer J, Warshel A (1999) J Phys Chem B 103:884–892. doi:10.1021/jp983699s
32. Ghosh A, Bansal M (2003) A glossary of DNA structures from A to Z. Acta Crystallogr D Biol Crystallogr 59:620–626. doi:10.1107/S0907444903003251
33. Dickerson RE, Drew HR, Conner BN, Wing RM, Fratini AV, Kopka ML (1982) The anatomy of A-, B-, and Z-DNA. Science 216:475–485. doi:10.1126/science.7071593
34. Sinden RR (1994) DNA structure and function. Academic, New York, pp 179–216
35. Basham B, Schroth GP, Ho PS (1995) Proc Natl Acad Sci USA 92:6464–6468. doi:10.1073/pnas.92.14.6464
36. Leslie AG, Arnott S, Chandrasekaran R, Ratliff RL (1980) J Mol Biol 143:49–72. doi:10.1016/0022-2836(80)90124-2
37. Rich A, Zhang S (2003) Nat Rev Genet 4:566–572. doi:10.1038/nrg1115
38. Egli M, Gessner RV (1995) Proc Natl Acad Sci USA 92:180–184. doi:10.1073/pnas.92.1.180