



# B-DNA Structure and Stability: The Role of Nucleotide Composition and Order

Celine Nieuwland,<sup>[a]</sup> Trevor A. Hamlin,<sup>[a]</sup> Célia Fonseca Guerra,<sup>\*,[a, b]</sup> Giampaolo Barone,<sup>\*,[c]</sup> and F. Matthias Bickelhaupt<sup>\*,[a, d]</sup>

We have quantum chemically analyzed the influence of nucleotide composition and sequence (that is, order) on the stability of double-stranded B-DNA triplets in aqueous solution. To this end, we have investigated the structure and bonding of all 32 possible DNA duplexes with Watson–Crick base pairing, using dispersion-corrected DFT at the BLYP-D3(BJ)/TZ2P level and COSMO for simulating aqueous solvation. We find enhanced stabilities for duplexes possessing a higher GC base pair content. Our activation strain analyses unexpectedly identify the loss of stacking interactions within individual

strands as a destabilizing factor in the duplex formation, in addition to the better-known effects of partial desolvation. Furthermore, we show that the sequence-dependent differences in the interaction energy for duplexes of the same overall base pair composition result from the so-called “diagonal interactions” or “cross terms”. Whether cross terms are stabilizing or destabilizing depends on the nature of the electrostatic interaction between polar functional groups in the pertinent nucleobases.

## Introduction

The genetic information encoded in the DNA of living organisms is the foundation of life.<sup>[1]</sup> Therefore, understanding the dynamic geometry and stability of DNA represents a thriving field.<sup>[2]</sup> By now it is well-known that the structure of the

double-stranded helix is strongly affected by factors such as the nucleotide composition, solvation, the presence of counterions, and interaction with proteins and small molecules (e.g., drugs).<sup>[3]</sup> Quantum chemical calculations are often employed to assess and quantify these variables individually and pinpoint their influence on the structure and stability of DNA molecules.<sup>[4]</sup> Many of these studies report analyses of simplified DNA model systems. Naturally, one begins with the study of structure and stability of the Watson–Crick DNA base pairs<sup>[4a–d]</sup> adenine–thymine (A–T) and guanine–cytosine (G–C), mismatches,<sup>[4a,c,d]</sup> DNA single strands,<sup>[4e,f]</sup> or double-stranded doublets (i.e., two stacked base pairs).<sup>[4c–f]</sup> The impact and importance of this work cannot be overstated, but quantum chemical computational studies of more realistic DNA models are necessary to see how these fundamental observations hold when all structural and environmental components are combined.<sup>[5]</sup> Density Functional Theory (DFT) modeling of triplet DNA helices of all-AT or all-GC base pairs with a neutralized sugar-phosphate backbone, immersed in a continuum type dielectric medium, provides geometries in agreement with the experimental B-DNA geometry and with accurate binding energies.<sup>[5a,f]</sup> More recently, this work was extended towards the double-stranded DNA pentamers d(A)<sub>5</sub>-d(T)<sub>5</sub> and d(G)<sub>5</sub>-d(C)<sub>5</sub> and the duplex Dickerson Dodecamer sequence d-(CGCGAATTCGCG).<sup>[5b]</sup> Different structural behavior and stabilities were observed for the d(A)<sub>5</sub>-d(T)<sub>5</sub> and d(G)<sub>5</sub>-d(C)<sub>5</sub> helices, which the authors ascribed to the presence of an extra hydrogen bond between the G–C base pairs. Nucleotide composition dependency was also observed in previous work of our group that showed that the stability of double-stranded DNA doublets increases with the G–C base-pair content, thus with the number of hydrogen bonds.<sup>[4f]</sup> In addition, an effect of the nucleotide sequence (i.e., order) on the doublet stability was observed. This indicates that there are sequence-dependent interactions that need to be understood to accurately

[a] C. Nieuwland, Dr. T. A. Hamlin, Prof. Dr. C. Fonseca Guerra, Prof. Dr. F. M. Bickelhaupt  
Department of Theoretical Chemistry,  
Amsterdam Institute of Molecular and Life Sciences (AIMMS),  
Amsterdam Center for Multiscale Modeling (ACMM)  
Vrije Universiteit Amsterdam  
De Boelelaan 1083  
1081 HV Amsterdam (The Netherlands)  
E-mail: c.fonseca Guerra@vu.nl  
f.m.bickelhaupt@vu.nl

[b] Prof. Dr. C. Fonseca Guerra  
Leiden Institute of Chemistry  
Gorlaeus Laboratories  
Leiden University  
Einsteinweg 55  
2300 CC Leiden (The Netherlands)

[c] Prof. Dr. G. Barone  
Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche  
Università degli Studi di Palermo  
Viale delle Scienze, Edificio 17  
90128 Palermo (Italy)  
E-mail: giampaolo.barone@unipa.it

[d] Prof. Dr. F. M. Bickelhaupt  
Institute of Molecules and Materials  
Radboud University Nijmegen  
Heyendaalseweg 135  
6525 AJ Nijmegen (The Netherlands)

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/open.202100231>

Part of the Chemistry Europe joint Special Collection on Quantum Bioinorganic Chemistry

© 2022 The Authors. Published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

predict the stabilities of nucleic acids. The reason for this is that the DNA double helix comprises not only hydrogen bonds between complementary bases, but also interactions between stacked bases and between bases that are diagonal to each other, that is, on opposite sides and in different layers.<sup>[4d]</sup>

In this work, we report an extensive analysis of intra-DNA interactions using dispersion-corrected DFT (DFT-D) computations of more realistic DNA structures: trideoxyribonucleoside diphosphate double strands (i.e., triplets, see Figure 1a). We consider double-stranded B-DNA triplets with a charged backbone ( $d(\text{DNA})^{-4}$ ) and a backbone neutralized by either  $\text{H}^+$  ( $d(\text{DNA})^{\text{H}^+}$ ) or  $\text{Na}^+$  ( $d(\text{DNA})^{\text{Na}^+}$ ) counterions. All 32 possible sequences of duplex DNA triplets were examined, considering Watson–Crick base pairing (Figure 1b). In the example of Figure 1a, the  $d(\text{CGG})^{\text{H}^+}$  duplex consists of the  $(\text{CGG})^{\text{H}^+}$  and  $(\text{CCG})^{\text{H}^+}$  complementary single strands with a CGG and CCG sequence in 5'-to-3' notation, respectively. This definition for the single and double strands differs slightly from our previous work,<sup>[4f]</sup> but we adjusted the notation to arrive at a more clear and compact notation for the triplets in this work.

First, we justify the accuracy of the DFT description of the DNA model systems by comparison of calculated geometrical parameters with available experimental X-ray data, after which the effect of the nucleotide composition, backbone neutralization, and nucleotide order on the DNA stability is examined using the activation strain model (ASM)<sup>[6]</sup> of reactivity and bonding, in combination with our canonical energy decomposition analysis (EDA).<sup>[7]</sup> Finally, a structural decomposition of the DNA molecules allows us to delineate the importance of the interactions between structural components to the overall stability of the double helix.

Detailed analyses of the structure and stability of single- and double-stranded DNA triplets is relevant, among others, because the 64 possible sequences of three bases in a single DNA (or RNA) strand constitute the codons, that is, the letters

or units of the genetic code that direct protein synthesis (note that the 64 possible single strands can form 32 unique Watson–Crick duplexes).<sup>[1]</sup> Furthermore, understanding the physical principles of DNA molecules might provide new insights into gene expression and genome stability. This work demonstrates that a detailed analysis of structural and stability properties of biological macromolecules of significant size is nowadays feasible using highly accurate quantum chemical methods and leads to new insights.

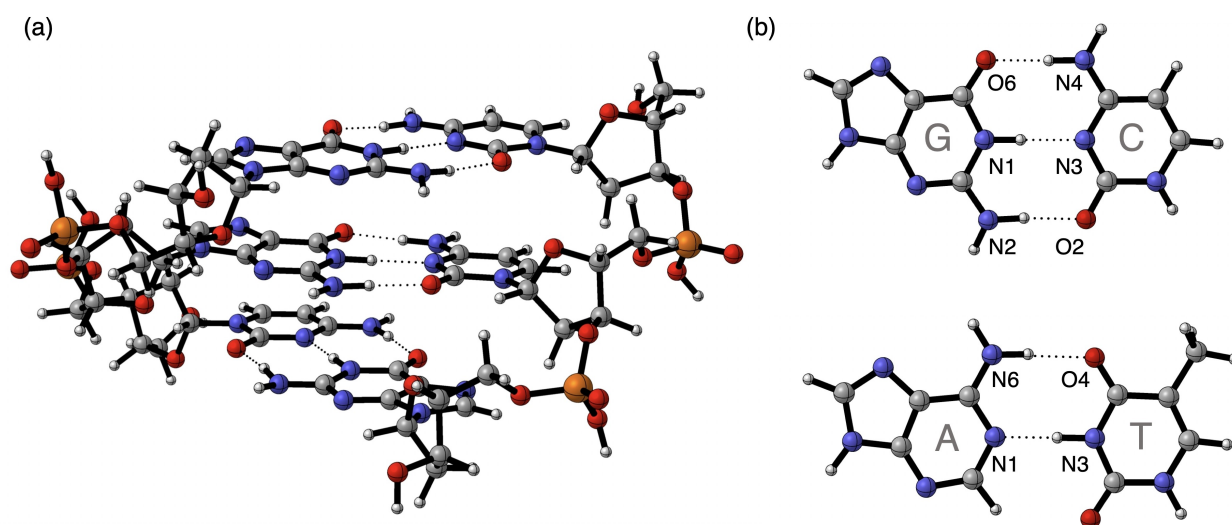
## Results and Discussion

### Validation of the Computational Method

All computations were performed at the BLYP-D3(BJ)/TZ2P level of theory using the Conductor-like Screening Model (COSMO) to simulate aqueous solvation (see also the Computational Details section). This level of theory has been proven to be accurate for our purpose, both in previous reports as well as in additional performance tests carried out in the present work.

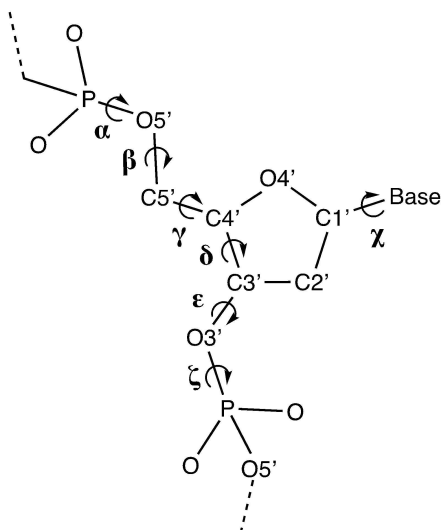
Various computational studies on DNA-based systems have demonstrated that BLYP-D3(BJ) yields reliable geometries and (hydrogen) bond energy trends that are in line with experimental data.<sup>[4c,f,8]</sup> BLYP-D3(BJ) furthermore outperforms more recently developed hybrid functionals, such as M06-2X, in terms of the accuracy/cost ratio in the description of non-covalent interactions.<sup>[9]</sup>

Herein, we have confirmed the reliability of our COSMO( $\text{H}_2\text{O}$ )-BLYP-D3(BJ)/TZ2P description for studying the 32  $d(\text{DNA})^{4-}$  model double-stranded triplets in the present work by comparing the computed geometries with available experimental X-ray diffraction data.<sup>[10]</sup> In our  $d(\text{DNA})^{4-}$  duplex model systems, the backbone neutralization is omitted, so that the overall structure has a net charge of  $-4$ . The optimized



**Figure 1.** (a) Geometry of  $d(\text{CGG})^{\text{H}^+}$ , as an example of the duplex B-DNA triplets studied in this work, optimized at the BLYP-D3(BJ)/TZ2P level using COSMO to simulate solvation in water. Hydrogen-bond interactions are depicted by dotted lines. Color code in ball-and-stick structures: H – white; C – grey; N – blue; O – red; P – yellow. (b) Watson–Crick DNA base pairs: guanine (G) – cytosine (C) and adenine (A) – thymine (T) with labeling of the hydrogen-bond front atoms.

geometries of the  $d(\text{DNA})^{4-}$  triplets were verified by comparison of the mean calculated sugar-phosphate backbone torsion angles (Scheme 1) to experimental mean values from B-DNA crystal structures available in the literature.<sup>[10]</sup> The mean calculated values are summarized in Table 1 (for details, including bond distances, see Tables S1 and S2 in the Supporting Information). In all cases, the obtained geometries are in very good agreement with the experimental data. Although in some cases the calculated mean backbone torsion angles of our relatively small dataset vary from the experimental mean values (up to  $22^\circ$  for the  $\beta$  angle), all calculated mean torsion angles fall well within the experimentally observed range of backbone torsion angles of B-DNA.<sup>[10]</sup> The use of the COSMO( $\text{H}_2\text{O}$ )-BLYP-D3(BJ)/TZ2P level of theory in our calculations is therefore justified since it is known from previous work that counterions, as are present under biological conditions and in our  $d(\text{DNA})^{\text{H}+}$  and  $d(\text{DNA})^{\text{Na}+}$  model systems, will further improve the agreement between the computed and experimental DNA geometries, in particular, backbone torsion angles.<sup>[5b,f]</sup>



**Scheme 1.** Definition of the sugar-phosphate backbone torsion angles in B-DNA structures.

Table 1. Mean backbone torsion angles (in $^\circ$ ; standard deviations in parentheses) of our 32 $d(\text{DNA})^{4-}$ structures from computation <sup>[a]</sup> compared to experiment. <sup>[b]</sup>		
	Computation	Experiment
alpha ( $\alpha$ )	-66 (2)	-62 (15)
beta ( $\beta$ )	-162 (9)	-184 (9)
gamma ( $\gamma$ )	50 (2)	48 (11)
delta ( $\delta$ )	141 (3)	128 (13)
epsilon ( $\epsilon$ )	166 (5)	184 (11)
zeta ( $\zeta$ )	-92 (3)	-95 (10)
chi ( $\chi$ )	-107 (4)	-102 (14)

[a] Our computations, performed at the COSMO( $\text{H}_2\text{O}$ )-BLYP-D3(BJ)/TZ2P level. See Scheme 1 for the definition of torsion angles; see the Supporting Information for Cartesian coordinates; [b] from X-ray crystal structures (resolution better than 1.9 Å) of 10 to 12 nucleotides long B-DNA oligodeoxynucleotides containing AT and GC base pair combinations.<sup>[10]</sup>

## B-DNA Stability: Effect of Nucleotide Composition

We have explored how the stability of B-DNA triplets depends on (i) the nucleotide composition, (ii) backbone neutralization by counterions, and (iii) the order in which the nucleotides occur in a strand by computing and analyzing the formation energy ( $\Delta E$ ) of the double helix from the two complementary single strands which is defined in Equation (1).

$$\Delta E = E_{\text{ds,aq}} - E_{\text{ss1,aq}} - E_{\text{ss2,aq}} \quad (1)$$

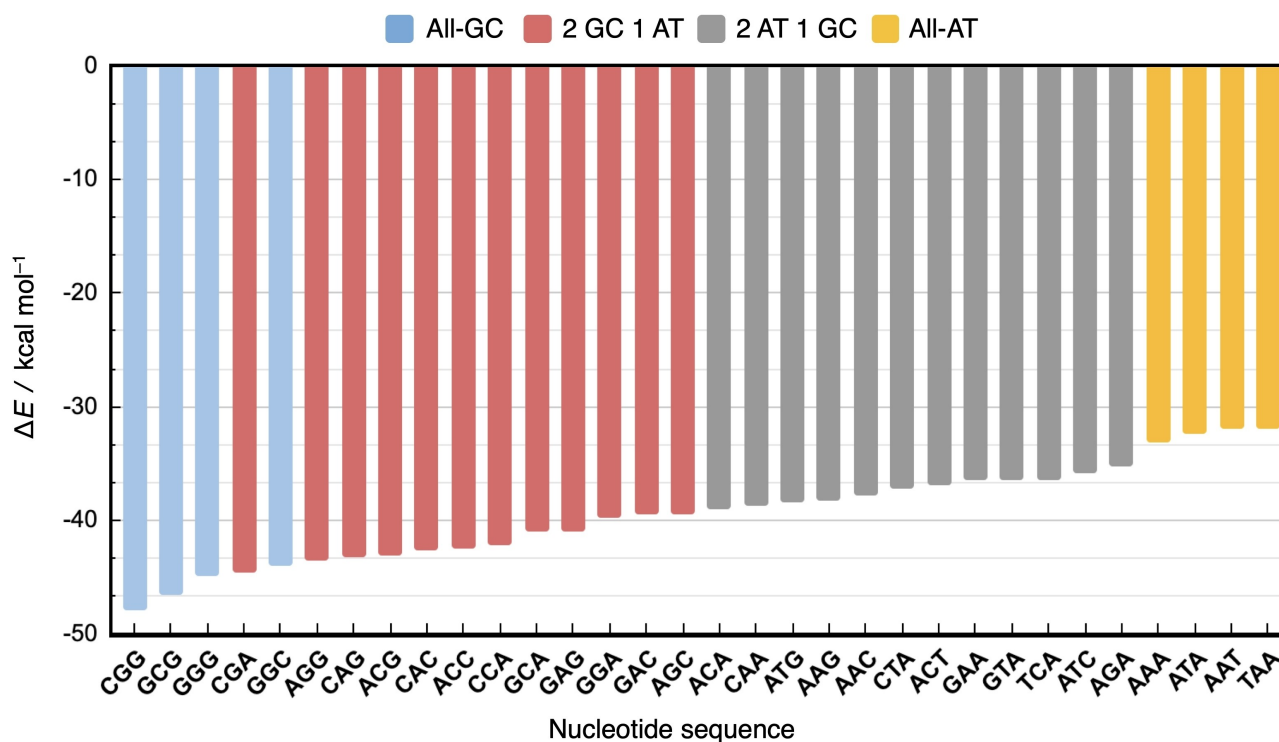
Here,  $E_{\text{ds,aq}}$  denotes the energy of the solvated DNA double strand (ds) in its equilibrium geometry.  $E_{\text{ss1,aq}}$  and  $E_{\text{ss2,aq}}$  correspond to the energies of the two separate solvated single strands (ss), each one in its own equilibrium geometry. The dissociated single strands are not rigid and can adopt a range of flexible orientations in solution. However, our analysis involves the initial stage after dissociation of the DNA duplex into the two constituent single strands. These separated single strands present a somewhat condensed structure that maximizes the intramolecular interactions (see Figure S1 for a graphical example).

Figure 2 shows our computed formation energies of the 32  $d(\text{DNA})^{\text{H}+}$  double strands involving non-ionized (neutral) phosphate groups. Numerical data can be found in Table S3 in the Supporting Information.

First, we examine the effect of the nucleotide composition, that is, the number of GC versus AT base pairs in the duplexes. The formation of all 32 unique duplexes from their complementary single strands is net stabilizing with formation energies  $\Delta E$  ranging from  $-32 \text{ kcal mol}^{-1}$  for the weakest duplex,  $d(\text{TAA})^{\text{H}+}$ , to  $-48 \text{ kcal mol}^{-1}$  for the most stable duplex,  $d(\text{CGG})^{\text{H}+}$ . The stabilization of  $\Delta E$  increases nearly proportionally as a function of the number of GC pairs. This trend is caused by the fact that the triply hydrogen-bonded G–C base pair is more stable than the only doubly hydrogen-bonded A–T base pair, in line with previous work.<sup>[4f,5b]</sup> Note, however, that for each nucleotide composition of the strands or, equivalently, Watson–Crick pair composition of the duplex (i.e., 0–3 G–C base pairs), there is still a substantial variation in the stabilities, depending on the order in which the base pairs occur, which can be as large as  $5 \text{ kcal mol}^{-1}$  in the case of duplexes containing two GC and one AT pair (see below).

The overall Gibbs free energy of formation,  $\Delta G$ , is less negative than the electronic energy of formation,  $\Delta E$ , mainly due to the destabilizing contributions of the zero-point vibrational energy,  $\Delta \text{ZPE}$ , and, predominantly, the entropy,  $T\Delta S$  (see below). The entropic penalty originates mainly from the increased overall rigidity when the relatively flexible single strands assemble to form the more rigid duplex. Importantly, the formation of duplexes remains spontaneous at room temperature (i.e.,  $\Delta G < 0$ ) in all 32 cases. Furthermore, the effect of the thermodynamic contributions (i.e.,  $\Delta G - \Delta E$ ) is estimated to be relatively constant, which suggests that trends in  $\Delta E$  are hardly affected and, by and large, carry over to  $\Delta G$ .

We have estimated this by explicitly carrying out the extremely costly computation of  $\Delta G$  for the most stable and for



**Figure 2.** Formation energy  $\Delta E$  (in kcal mol<sup>-1</sup>) of the 32 double-stranded DNA triplets d(DNA)<sup>H+</sup> (backbone neutralized by H<sup>+</sup>) from two complementary single strands, computed at the BLYP-D3(BJ)/TZ2P level, using COSMO to simulate solvation in water. The colors indicate sequences with the same relative number of GC and AT base pairs.

the least stable duplexes with a charged backbone, d(CGG)<sup>4-</sup> (187 atoms) and d(TAA)<sup>4-</sup> (190 atoms), respectively. The duplexes d(DNA)<sup>4-</sup> with charged backbones have been chosen because they are less stable than d(DNA)<sup>H+</sup> and d(DNA)<sup>Na+</sup> (see below), and therefore are in theory more likely to become unbound due to thermodynamic corrections. Thus, we have performed frequency calculations for d(CGG)<sup>4-</sup>, d(TAA)<sup>4-</sup>, as well as for the respective single strands, to obtain the vibrational states in the pertinent partition functions. d(CGG)<sup>4-</sup> is formed from the two complementary single strands (CGG)<sup>2-</sup> and (CCG)<sup>2-</sup>, whereas d(TAA)<sup>4-</sup> is constructed from the strands (TAA)<sup>2-</sup> and (TTA)<sup>2-</sup>. The calculated  $\Delta G$  values are  $-22.3$  and  $-6.9$  kcal mol<sup>-1</sup> for d(CGG)<sup>4-</sup> and d(TAA)<sup>4-</sup>, respectively. The corresponding  $\Delta E$  values are found to be  $-44.8$  and  $-28.8$  kcal mol<sup>-1</sup> (see Supporting Information Table S3). Thus, there is a destabilizing and relatively constant  $\Delta ZPE$  and entropic  $T\Delta S$  contribution in the double-helix formation of 22 to 23 kcal mol<sup>-1</sup>, that applies to all DNA triplets. Therefore, the sequence-dependent trend in duplex stability primarily depends on the electronic energy of formation,  $\Delta E$ .

### B-DNA Stability: Effect of Counterions

Our analyses focus mainly on neutral DNA triplets d(DNA)<sup>H+</sup> with non-ionized phosphate groups (see for example, Figure 2) because, as further explorations show, they are computationally more stable than ionized (anionic) triplets or triplets with

sodium counterions while displaying near-, although not fully, identical trends in stability as a function of their sequences. Thus, DNA model duplexes with anionic sugar-phosphate backbones (d(DNA)<sup>4-</sup>), those with neutral sugar-phosphate backbones that we mainly focus on (d(DNA)<sup>H+</sup>, that is, neutralized with H<sup>+</sup> counter ions), and those with sugar-phosphate backbones neutralized by Na<sup>+</sup> counter ions (d(DNA)<sup>Na+</sup>) show very similar trends in stability as function of the nucleotide composition, as can be seen by comparing Figure 2 and Supporting Information Figures S2 and S3.

Although backbone neutralization has little effect on stability trends as a function of base sequence, it has an important impact on the *absolute* B-DNA stability. The formation energy  $\Delta E$  becomes increasingly stabilizing along the order d(DNA)<sup>4-</sup> < d(DNA)<sup>H+</sup> < d(DNA)<sup>Na+</sup>. For example, along d(CGG)<sup>4-</sup>, d(CGG)<sup>H+</sup>, and d(CGG)<sup>Na+</sup>, the formation energy  $\Delta E$  becomes increasingly stabilizing from  $-44.8$  to  $-47.9$  to  $-51.8$  kcal mol<sup>-1</sup>, respectively (see Table S3). The reason for the enhanced stability upon backbone neutralization is, not unexpectedly, a more favorable electrostatic interaction (see the difference in  $\Delta V_{\text{elstat}}$  of d(CGG)<sup>H+</sup> versus d(CGG)<sup>4-</sup> in Table S4).

### B-DNA Stability: Effect of Nucleotide Order

As pointed out earlier, the stability of a duplex does not only depend on the base pair composition but also on the order in which the base pairs occur. In the following part, we show that

sequence-dependent variations in the stability of the duplex assembling from two complementary single strands arise mainly from three competing terms: (i) the desolvation; (ii) the deformation strain; and (iii) diagonal interactions, that is, cross terms. The meaning and implications of these factors will become clear in a moment, but the individual terms have a counteracting effect in such a way that the most stable DNA sequence is obtained at the optimum compromise of these terms. We first identify the general trends in these three terms, after which we explain these trends by an in-depth analysis of each term. The nucleotide order effects are subtle and sometimes hard to foresee. Nevertheless, we can provide some useful guidelines for predicting the stabilities of oligonucleotides at the end of this work.

To examine the effect of the nucleotide order on the duplex stability, let us focus on the two extreme cases in Figure 2, that is, the all-GC (blue) and all-AT (yellow) d(DNA)<sup>H+</sup> duplexes. Among the former, d(CGG)<sup>H+</sup> is the most stable duplex, followed by d(GCG)<sup>H+</sup>, d(GGG)<sup>H+</sup>, and d(GGC)<sup>H+</sup>. Among the latter, that is, the all-AT duplexes, d(AAA)<sup>H+</sup> is the most stable sequence, followed by d(ATA)<sup>H+</sup>, d(AAT)<sup>H+</sup>, and d(TAA)<sup>H+</sup>, respectively. To pinpoint the origin of the varying stability among these all-GC and AT sequences, the duplex formation energy,  $\Delta E$ , was partitioned using the activation strain model (ASM)<sup>[6]</sup> as formulated by Equation (2) and illustrated by Figure 3.

$$\Delta E = -\Delta E_{\text{solv,ss1+ss2}} + \Delta E_{\text{strain}} + \Delta E_{\text{int}} + \Delta E_{\text{solv,ds}} \quad (2)$$

The desolvation energy (i.e., the negative of the solvation energy) of the two complementary single strands ( $-\Delta E_{\text{solv,ss1+ss2}}$ ) and the solvation energy of the duplex ( $\Delta E_{\text{solv,ds}}$ ) are computed by taking the energy difference going from the solvated species

in their solution-phase equilibrium geometry to the same species in the gas phase but still in their solution-phase equilibrium geometry (see Equations (3) and (4)).

$$-\Delta E_{\text{solv,ss1+ss2}} = E_{\text{ss1,gas}} + E_{\text{ss2,gas}} - E_{\text{ss1,aq}} - E_{\text{ss2,aq}} \quad (3)$$

$$\Delta E_{\text{solv,ds}} = E_{\text{ds,aq}} - E_{\text{ds,gas}} \quad (4)$$

The subscript 'aq' refers to computations in aqueous solution whereas the subscript 'gas' refers to gas-phase single-point calculations on the solution-phase equilibrium geometry. Thus,  $E_{\text{ss1,gas}}$ ,  $E_{\text{ss2,gas}}$ , and  $E_{\text{ds,gas}}$  denote the gas-phase energies of the two single DNA strands and the DNA duplex, each in their equilibrium geometry in aqueous solution.

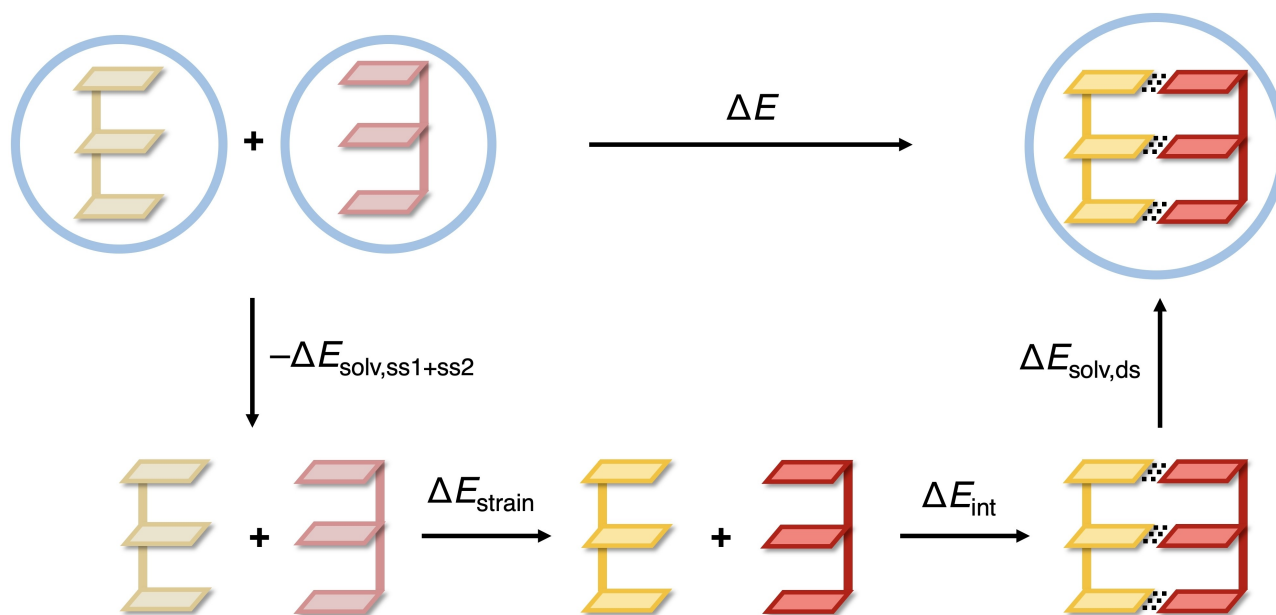
The strain energy  $\Delta E_{\text{strain}}$  is the energy required to deform the two complementary single strands, in the gas phase, from their equilibrium geometry in aqueous solution, with the energy  $E_{\text{ss1,gas}} + E_{\text{ss2,gas}}$  to the structure they adopt in the DNA duplex in its equilibrium geometry in aqueous solution, with the energy  $E_{\text{ss1',gas}} + E_{\text{ss2',gas}}$  (Equation (5)).

$$\Delta E_{\text{strain}} = E_{\text{ss1',gas}} + E_{\text{ss2',gas}} - E_{\text{ss1,gas}} - E_{\text{ss2,gas}} \quad (5)$$

The interaction energy  $\Delta E_{\text{int}}$  corresponds to the net stabilization associated with combining the two deformed single strands, again in the absence of any solvent, that is, with an energy of  $E_{\text{ss1',gas}} + E_{\text{ss2',gas}}$  to form the double helix, with an energy of  $E_{\text{ds,gas}}$  (Equation (6)).

$$\Delta E_{\text{int}} = E_{\text{ds,gas}} - E_{\text{ss1',gas}} - E_{\text{ss2',gas}} \quad (6)$$

Note that in our energy decomposition as outlined in Figure 3 and Equation (2), in particular, the interaction energy



**Figure 3.** Partitioning of the formation energy ( $\Delta E$ ) of the DNA duplex assembling from two complementary single strands (blue circles represent aqueous solvent medium).

$\Delta E_{\text{intr}}$  is based on charge-neutralized DNA duplexes with  $\text{H}^+$  counterions ( $\text{d}(\text{DNA})^{\text{H}^+}$ ) that provide stabilities, and trends therein, nicely match those occurring under physiological conditions, that is, in the presence of sodium counterions (see above). Importantly, our analysis is free of spurious Coulomb repulsion that would occur in the gas phase for non-neutralized  $\text{d}(\text{DNA})^{4-}$  model systems (compare  $\Delta V_{\text{elstat}}$  for  $\text{d}(\text{CGG})^{\text{H}^+}$  and  $\text{d}(\text{CGG})^{4-}$  in Table S4).

The partitioning of the formation energy  $\Delta E$  is reported in Table 2 for the all-GC and all-AT  $\text{d}(\text{DNA})^{\text{H}^+}$  sequences (see Figure S4 for a graphical representation). The magnitudes of the energy terms are larger for the GC sequences than for the AT sequences, which can be attributed to the presence of additional hydrogen bonds in the former. Also, for some energy terms, we observe a more pronounced sequence-dependent variation for the GC sequences (see below). From Table 2, it follows that, in all cases,  $\Delta E$  is the result of a large, stabilizing interaction energy, counteracted by smaller, but still significant, destabilizing contributions from deformation strain and solvent effects.

The  $\Delta E_{\text{strain}}$  term varies only slightly for the different  $\text{d}(\text{DNA})^{\text{H}^+}$  sequences. The least destabilizing strain energy is observed for the sequences  $\text{d}(\text{GCG})^{\text{H}^+}$  and  $\text{d}(\text{ATA})^{\text{H}^+}$ . This can be ascribed to the fact that the associated single strands are characterized by an alternating arrangement of purine and the smaller pyrimidine bases which leads to less stabilizing internal stacking interactions than in the case of single strands involving stacks of two or three purines, for example, in the case of  $\text{d}(\text{CGG})^{\text{H}^+}$  (see Figure 1a) and  $\text{d}(\text{AAA})^{\text{H}^+}$ , that go with larger mutual surface overlap and thus stronger stacking interactions. Consequently, the geometrical deformation required to form the double helix in the case of  $\text{d}(\text{GCG})^{\text{H}^+}$  and  $\text{d}(\text{ATA})^{\text{H}^+}$  also goes with a smaller loss of favorable stacking interaction and, thus, less strain  $\Delta E_{\text{strain}}$  than in the case of  $\text{d}(\text{CGG})^{\text{H}^+}$  and  $\text{d}(\text{AAA})^{\text{H}^+}$ .

The solvation of the two single strands ( $\Delta E_{\text{solv,ss1+ss2}}$ ) is more stabilizing than the solvation of the corresponding double strand ( $\Delta E_{\text{solv,ds}}$ ), as can be seen from the data in Table 2. This is because the hydrogen-bond front atoms need to be partly desolvated to facilitate the complementary base pairing. This partial desolvation is the reason for the overall destabilizing effect of the solvent represented by  $\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$  in Table 2. Furthermore, there is only little variation in  $\Delta E_{\text{solv,ds}}$

among the all-AT sequences while there is a more pronounced sequence dependence in this term for the all-GC duplexes (see below).

Finally, a trend opposite to  $\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$  was found for the interaction energy  $\Delta E_{\text{intr}}$  meaning that when the overall desolvation is strong, the interaction energy is also strong as is the case for the CGG and GGG, and TAA and AAA sequences. This can be understood with the insight that the interaction with the aqueous solvent and between the complementary DNA bases relies both on hydrogen-bond interactions. Therefore  $\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$  and  $\Delta E_{\text{intr}}$  should yield the same trend, but with an opposite sign. A more detailed analysis of the interaction energy  $\Delta E_{\text{intr}}$  can be found in Table S4.<sup>[7]</sup>

### Interaction Analysis: Decomposition of the Solvation Energy

Next, we address the origin of the computed solvent effects  $\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$  and the aforementioned trends therein. To this end, we decompose the solvation energy  $\Delta E_{\text{solv}} (= E_{\text{aq}} - E_{\text{gas}})$  into the components that are associated with two different physical phenomena, namely, the cavitation energy ( $E_{\text{solv,cav}}$ ) and the solute–solvent interaction ( $E_{\text{solv,int}}$ ) (see Equation (7)).

$$\Delta E_{\text{solv}} = E_{\text{solv,cav}} + E_{\text{solv,int}} \quad (7)$$

The former,  $E_{\text{solv,cav}}$  is the energy associated with creating an empty cavity in the solvent to make room for the solute molecules, that is, the single and double DNA strands. Cavitation disrupts the local (fluxional) solvent structure and is accompanied by loss of intermolecular interactions between solvent molecules. In the COSMO model that we use in our computations, this is simulated by a destabilizing energy term, associated with creating a cavity in a dielectric medium of the right size and shape to host the solute molecules based on effective radii of the solvent and the atoms in the solute (for details, see Ref. [11]). The solute–solvent interaction  $E_{\text{solv,int}}$  results from the interaction of the solute molecule in the cavity with the solvent. In COSMO, this comes down to the interaction of the fully quantum chemically described solute molecules that generate, and electrostatically interact with, the mirror charges generated on the surface of the cavity in the dielectric continuum.<sup>[11]</sup> Note that, also in our COSMO simulation, the

**Table 2.** Partitioning of the formation energy  $\Delta E$  (in  $\text{kcal mol}^{-1}$ ) for the  $\text{d}(\text{DNA})^{\text{H}^+}$  duplex assembling from two complementary single strands for the all-GC and all-AT sequences.<sup>[a]</sup>

$\text{d}(\text{DNA})^{\text{H}^+}$	$\Delta E$	$-\Delta E_{\text{solv,ss1+ss2}}$	$\Delta E_{\text{strain}}$	$\Delta E_{\text{intr}}$	$\Delta E_{\text{solv,ds}}$	$\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$
$\text{d}(\text{CGG})^{\text{H}^+}$	−47.9	161.8	19.1	−118.8	−110.0	51.8
$\text{d}(\text{GCG})^{\text{H}^+}$	−46.6	146.4	16.6	−101.9	−107.7	38.7
$\text{d}(\text{GGG})^{\text{H}^+}$	−44.9	168.7	21.9	−117.3	−118.2	50.5
$\text{d}(\text{GGC})^{\text{H}^+}$	−44.0	153.3	17.8	−101.5	−113.6	39.7
$\text{d}(\text{AAA})^{\text{H}^+}$	−33.2	122.9	7.8	−64.6	−99.3	23.6
$\text{d}(\text{ATA})^{\text{H}^+}$	−32.4	121.5	5.3	−59.4	−99.8	21.7
$\text{d}(\text{AAT})^{\text{H}^+}$	−32.0	121.0	7.5	−60.8	−99.7	21.3
$\text{d}(\text{TAA})^{\text{H}^+}$	−32.0	123.1	7.2	−63.1	−99.1	24.0

[a] Computed at the BLYP-D3(BJ)/TZ2P level using COSMO to simulate solvation in water. See also Equation (2).

solute–solvent interaction induces a change in the density of the solute, although no charge is transferred between solute and solvent. The results of the solvation energy decomposition of the all-GC and all-AT duplexes, based on Equation (7), are collected in Table 3.

Solute–solvent interactions are substantially stabilizing for all single and double DNA strands, whereas cavitation terms are smaller and destabilizing, as expected. The solute–solvent interaction term ( $E_{\text{solv,int}}$ ) becomes less stabilizing going from the single-strands to the double-strand as the duplex formation requires partial desolvation of the nucleobases, where stabilizing interactions with aqueous continuum are exchanged for the Watson–Crick hydrogen bonds. The cavitation term  $E_{\text{solv,cav}}$  becomes only slightly less destabilizing going from the two single strands to the double strand, which is associated with a decrease in cavity size within the continuum solvent. The contraction of the cavity is the result of the hydrogen-bond interactions between the DNA base pairs which allows them to approach each other more closely than the sum of their van der Waals radii.<sup>[1b]</sup> In Watson–Crick base pairs, this hydrogen-bond contraction amounts to approximately 0.5 Å. Overall, the minor decrease of  $E_{\text{solv,cav}}$  does not make up for the significant loss in solute–solvent interactions, leaving the total solvent effect as a destabilizing factor in the duplex formation

The dependence of the overall solvent effect ( $\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$ ) on the order in which the nucleotides occur results from the variation in solute–solvent interactions, as  $E_{\text{solv,cav}}$  hardly varies among the all-GC and all-AT sequences (Table 3). As the formation of the Watson–Crick hydrogen bonds shields the interior of the duplex, the sequence dependence of the solvent effect originates primarily from the difference in solute–solvent interactions of the single strands as supported by the more significant variation of  $E_{\text{solv,int}}$  for the two single strands compared to the double strands in Table 3. Remarkably, there is only little variation in  $\Delta E_{\text{solv,ss1+ss2}}$  among the all-AT sequences, while there is a clear sequence dependence in this term for the all-GC duplexes. The reason for this is that the solvation energies of the nucleobases G and C differ more than the solvation energies of A and T, leading to more pronounced sequence-dependent differences for the all-GC single-strands (see Figure S5a). In addition to the larger G versus C solvation energy difference, our computations identify that the  $\text{NH}_2$

group of G is positioned such that it can form an intra-strand hydrogen bond with the deoxyribose oxygen atom of the backbone to which it is attached when located in the first or second position (5'-to-3' notation) of a triplet single strand (see Figure S5b). The formation of internal hydrogen bonds reduces the desolvation energy ( $-\Delta E_{\text{solv,ss1+ss2}}$ ) of all single-strands with a G in position 1 or 2 upon forming the duplex (see Table S5). However, as the  $(\text{CCC})^{\text{H}+}$  and  $(\text{CCG})^{\text{H}+}$  single strands do not contain such guanines and therefore do not form intra-strand hydrogen bonds, this leads to relatively higher desolvation energies of these strands. As the  $(\text{CCC})^{\text{H}+}$  and  $(\text{CCG})^{\text{H}+}$  single strands are part of the  $\text{d}(\text{GGG})^{\text{H}+}$  and  $\text{d}(\text{CCG})^{\text{H}+}$  duplexes, respectively, they lead to more destabilizing  $\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$  values for these duplexes (Table 2). The position of the  $\text{NH}_2$  group in adenine cannot facilitate hydrogen bonds with the sugar-phosphate backbone, which also explains the lack of sequence dependency in  $\Delta E_{\text{solv,ds}} - \Delta E_{\text{solv,ss1+ss2}}$  for the all-AT duplexes.

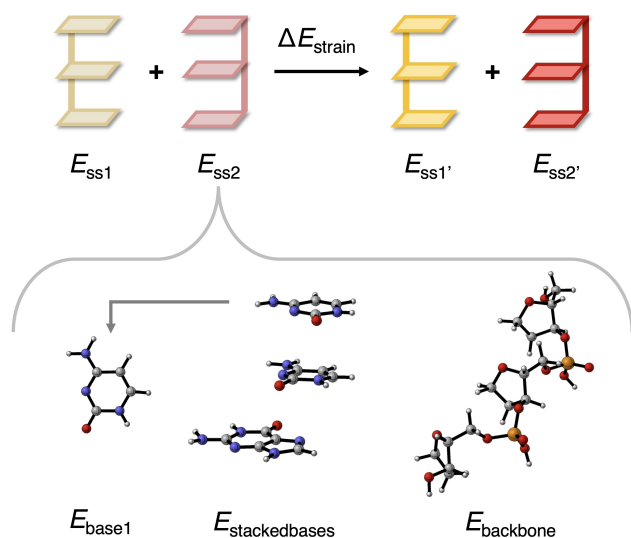
### Interaction Analysis: Decomposition of the Strain Energy

The strain energy  $\Delta E_{\text{strain}}$  of the single strands consists of the deformation of the sugar-phosphate backbone and the nucleobases due to the formation of the DNA double strand (see below). In this section, we quantify, for the various all-GC and all-AT sequences, which part of the deforming single strands contributes to what extent to the strain energy  $\Delta E_{\text{strain}}$  (see Table 2) upon duplex formation: (i) the sugar-phosphate backbone; (ii) the individual nucleobases; and (iii) the stacks of nucleobases. Using the COSMO-optimized geometries of the DNA double and single strands, the contributions were examined by decomposing the single strands, in the gas phase, into smaller segments, as illustrated in Figure 4. The resulting smaller fragments were kept in the exact same geometry as they had in the larger strand and the dissociated glycosidic C–N bonds were terminated by hydrogen atoms, yielding C–H and N–H bonds in the backbone and the nucleobase fragments, respectively. These terminating C–H and N–H bonds have the same bond and dihedral angles as the original glycosidic bond and a fixed bond length of 1.03 (N–H) or 1.07 Å (C–H).

**Table 3.** Partitioning of the solvation energy  $\Delta E_{\text{solv}}$  (in kcal mol<sup>-1</sup>) of the double strand (ds) compared to the two complementary single strands (ss1 + ss2) for the all-GC and all-AT d(DNA)<sup>H+</sup> duplexes.<sup>[a]</sup>

d(DNA) <sup>H+</sup>	ss1 + ss2			ds		
	$\Delta E_{\text{solv}}$	$E_{\text{solv,cav}}$	$E_{\text{solv,int}}$	$\Delta E_{\text{solv}}$	$E_{\text{solv,cav}}$	$E_{\text{solv,int}}$
d(CGG) <sup>H+</sup>	−161.8	7.8	−169.6	−110.0	6.8	−116.8
d(GCG) <sup>H+</sup>	−146.4	7.6	−154.0	−107.7	6.8	−114.5
d(GGG) <sup>H+</sup>	−168.7	7.5	−176.2	−118.2	6.8	−124.9
d(GGC) <sup>H+</sup>	−153.3	7.5	−160.7	−113.5	6.8	−120.3
d(AAA) <sup>H+</sup>	−122.9	7.6	−130.6	−99.3	6.9	−106.2
d(ATA) <sup>H+</sup>	−121.5	7.7	−129.3	−99.8	7.0	−106.8
d(AAT) <sup>H+</sup>	−121.0	7.6	−128.6	−99.7	6.9	−106.6
d(TAA) <sup>H+</sup>	−123.1	7.8	−130.9	−99.1	7.1	−106.2

[a] Computed at the BLYP-D3(BJ)/TZ2P level using COSMO to simulate solvation in water.



**Figure 4.** Structural decomposition of  $(\text{DNA})^{\text{H}+}$  to evaluate the contributions of the nucleobases, base stacking, and sugar-phosphate backbone to the total strain energy  $\Delta E_{\text{strain}}$ . Color code in ball-and-stick structures: H – white; C – grey; N – blue; O – red; P – yellow.

The contribution to  $\Delta E_{\text{strain}}$  stemming from the deformation of the sugar-phosphate backbone ( $\Delta E_{\text{strain\_backbone}}$ , Equation (8)) is estimated by the difference between the energies of the backbone fragments in the separate single strands ( $E_{\text{backbone\_ss1}}$  and  $E_{\text{backbone\_ss2}}$ , see Figure 4) and the energies of the backbone fragments in the duplex geometry ( $E_{\text{backbone\_ss1'}}$  and  $E_{\text{backbone\_ss2'}}$ ).

$$\Delta E_{\text{strain\_backbone}} = E_{\text{backbone\_ss1'}} + E_{\text{backbone\_ss2'}} - E_{\text{backbone\_ss1}} - E_{\text{backbone\_ss2}} \quad (8)$$

The contribution to  $\Delta E_{\text{strain}}$  stemming from the individual nucleobases upon forming the DNA duplex from the two complementary strands ( $\Delta E_{\text{strain\_bases}}$ ) is estimated using Equation (9). Here,  $E_{\text{base}}$  denotes the energy of a separate base in the geometry of the separate single strands ( $E_{\text{base\_ss1}}$  and  $E_{\text{base\_ss2}}$ , see Figure 4) or in the deformed geometry within the DNA double strand ( $E_{\text{base\_ss1'}}$  and  $E_{\text{base\_ss2'}}$ ). So, Equation (9) sums up, for each strand, the deformation of the three individual bases within each strand.

$$\Delta E_{\text{strain\_bases}} = \sum_{i=1}^3 (E_{\text{base\_ss1}'} - E_{\text{base\_ss2}})_i + (E_{\text{base\_ss2}'} - E_{\text{base\_ss2}})_i \quad (9)$$

Finally, the contribution to  $\Delta E_{\text{strain}}$  stemming from the change in base stacking interactions per strand is estimated by  $\Delta E_{\text{strain\_stacking}}$  as formulated in Equation (10). Here,  $\Delta E_{\text{strain\_stacking}}$  is the difference in stacking energy going from the DNA single strands ( $\Delta E_{\text{stacking\_ss1}}$  and  $\Delta E_{\text{stacking\_ss2}}$ ) to the respective strands in the duplex ( $\Delta E_{\text{stacking\_ss1'}}$  and  $\Delta E_{\text{stacking\_ss2'}}$ ). The stacking energy  $\Delta E_{\text{stacking}}$  of an individual strand is formulated as the energy of the stack of bases ( $E_{\text{stackedbases}}$ , see Figure 4) minus the energy of the three separate bases ( $E_{\text{base1-3}}$ ) in the geometry within the respective stack (see Equation (11)).

$$\Delta E_{\text{strain\_stacking}} = \Delta E_{\text{stacking\_ss1'}} + \Delta E_{\text{stacking\_ss2'}} - \Delta E_{\text{stacking\_ss1}} - \Delta E_{\text{stacking\_ss2}} \quad (10)$$

$$\Delta E_{\text{stacking}} = E_{\text{stackedbases}} - E_{\text{base1}} - E_{\text{base2}} - E_{\text{base3}} \quad (11)$$

The estimates of the three different contributions to the strain energy  $\Delta E_{\text{strain}}$  in the single strands upon duplex formation of the all-GC and all-AT sequences are collected in Table 4. Although the sum of  $\Delta E_{\text{strain\_backbone}}$ ,  $\Delta E_{\text{strain\_bases}}$ , and  $\Delta E_{\text{strain\_stacking}}$  covers most of the components that give rise to the activation strain of duplex formation, we stress that these terms do not add up exactly to the exact  $\Delta E_{\text{strain}}$  of the full duplexes, because of the fragmentation scheme which treats, in particular, the effects associated with the glycosidic bonds and the relative orientation of the backbone with regard to the nucleobases in an approximate manner.

$\Delta E_{\text{strain}}$  originates primarily from the deformation of the nucleobases and the loss of stacking interactions, while the backbone geometry deforms only slightly or even adopts a slightly more favorable geometry in the duplex. Interestingly, the values of  $\Delta E_{\text{strain\_backbone}}$  in Table 4 reflect that some of the backbone strain is relieved upon going from the geometry in the single strand to the hydrogen-bonded geometry in the DNA duplex. This shows that in the single strands, the equilibrium geometry is mainly determined by optimizing the base stacking interactions, which leads to a suboptimal

**Table 4.** Energy contributions<sup>[a]</sup> (in kcal mol<sup>-1</sup>) of the deformation of the sugar-phosphate backbone, base geometry, and base stacking to the exact strain energy  $\Delta E_{\text{strain}}$  for the  $(\text{DNA})^{\text{H}+}$  duplex assembling from the two complementary single strands for the all-GC and all-AT sequences.

$d(\text{DNA})^{\text{H}+}$	$\Delta E_{\text{strain\_backbone}}$	$\Delta E_{\text{strain\_bases}}$	$\Delta E_{\text{strain\_stacking}}$	Total <sup>[b]</sup>	$\Delta E_{\text{strain}}(\text{exact})$
$d(\text{CGG})^{\text{H}+}$	-1.2	9.2	6.8	14.8	19.1
$d(\text{GCG})^{\text{H}+}$	-2.5	9.0	4.6	11.1	16.6
$d(\text{GGG})^{\text{H}+}$	-3.3	8.9	8.9	14.5	21.9
$d(\text{GGC})^{\text{H}+}$	-3.4	8.8	3.5	8.9	17.8
$d(\text{AAA})^{\text{H}+}$	-2.3	5.1	3.1	5.9	7.8
$d(\text{ATA})^{\text{H}+}$	0.2	5.0	0.4	5.6	5.3
$d(\text{AAT})^{\text{H}+}$	0.6	5.1	1.7	7.4	7.5
$d(\text{TAA})^{\text{H}+}$	-1.0	4.9	2.4	6.3	7.2

[a] Computed at the BLYP-D3(BJ)/TZ2P level; [b] the sum of  $\Delta E_{\text{strain\_backbone}}$ ,  $\Delta E_{\text{strain\_bases}}$ , and  $\Delta E_{\text{strain\_stacking}}$ .



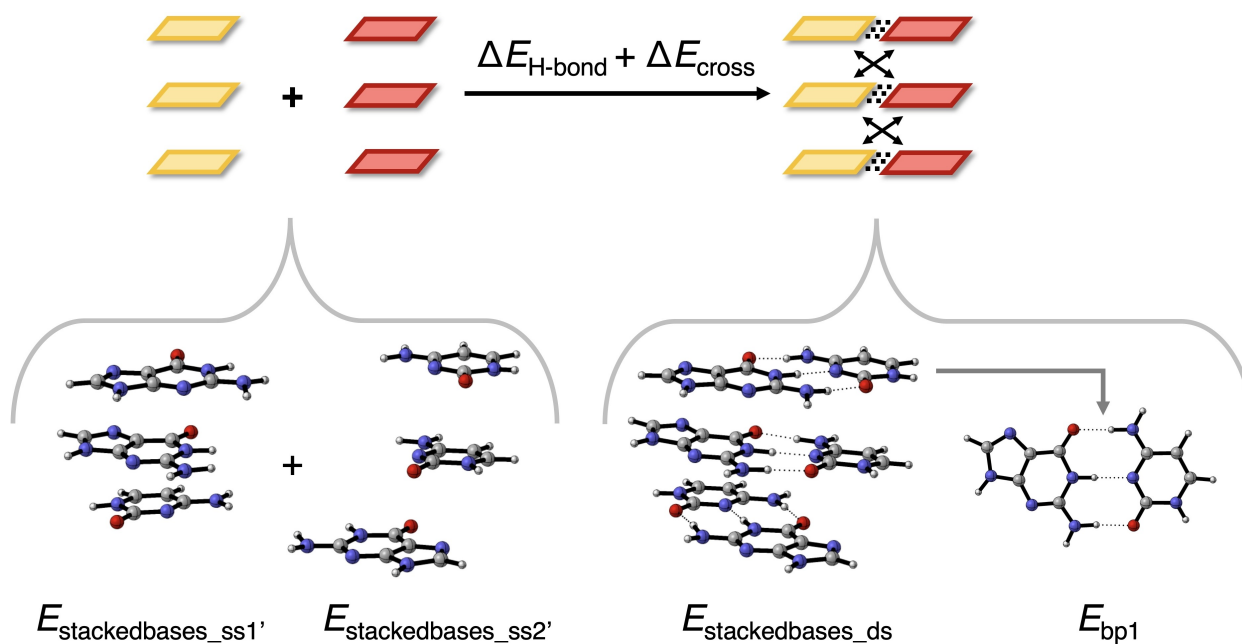
geometry for the sugar-phosphate backbone. At variance, in the DNA duplex, the equilibrium geometry is dominated by optimizing the relatively strong Watson–Crick hydrogen bonds which happen to match better with the intrinsic optimum of the sugar-phosphate backbone.

Thus, while base stacking is often regarded as the main contributor to duplex DNA stability,<sup>[12]</sup> we demonstrate here that, in fact, the base stacking becomes *less* favorable in the duplex geometry compared to the two separate single strands, as a destabilizing  $\Delta E_{\text{strain\_stacking}}$  was observed for all DNA sequences in Table 4. Important to realize here is that DNA duplexes do not assemble directly from individual unstacked nucleosides. Instead, driven by the formation of hydrogen bonds, the double helix is obtained from the separately synthesized complementary strands where stacking interactions are already present.<sup>[11a,13]</sup> Although base stacking contributes to the overall stability of DNA, based on the stacking interactions alone the formation of the duplex from two single-strands would be unfavored. Nevertheless, the duplex formation is ultimately promoted by the presence of other stabilizing interactions between the complementary strands, including hydrogen bonding. Due to the formation of the hydrogen bonds, the nucleobases also deform upon forming the DNA double strand.<sup>[14]</sup>  $\Delta E_{\text{strain\_bases}}$  is therefore destabilizing and amounts to approximately 9 kcal mol<sup>-1</sup> for the all-GC sequences, but only 5 kcal mol<sup>-1</sup> for the AT sequences. This difference is again because the G–C base pair forms three hydrogen bonds, while there are only two hydrogen bonds in the A–T base pair, which gives rise to more deformation in the former and less in the latter.

### Interaction Analysis: Decomposition of the Interaction Energy

Lastly, we analyze the interaction energy ( $\Delta E_{\text{int}}$ ) to identify the origin of the sequence dependence observed for the different all-GC and all-AT d(DNA)<sup>H+</sup> sequences (recall Table 2). Our analyses reveal that  $\Delta E_{\text{int}}$  consists of two major contributions: (i) the Watson–Crick hydrogen bonds; and (ii) the diagonal interactions between bases of two different, consecutively stacked Watson–Crick pairs. Interestingly, we find that the Watson–Crick hydrogen bonding interactions are sequence-independent. Therefore, the sequence dependence originates nearly exclusively from the diagonal interactions. Using the COSMO-optimized geometries of the d(DNA)<sup>H+</sup> duplexes, both interaction energy components were examined in the gas phase by decomposing the double strands into smaller segments, as illustrated in Figure 5. For this analysis, we employ nucleobase fragments without the sugar-phosphate backbone moiety. The consideration of only the nucleobases to describe the trends in total interaction energy is justified, as the interaction between the two sugar-phosphate backbone fragments was found to be negligible for the all-AT and all-GC sequences (see Table S6). The resulting smaller nucleobase fragments were kept in the exact same geometry as they had in the strand with backbone and the dissociated glycosidic C–N bonds were terminated by hydrogen atoms, yielding N–H bonds that have the same bond and dihedral angles as the original glycosidic bond and a fixed bond length of 1.03 Å.

As defined by Equation (12) and illustrated in Figure 5, the total interaction energy between two complementary stacks of bases, with energy  $E_{\text{stackedbases\_ss1}'}$  and  $E_{\text{stackedbases\_ss2}'}$ , upon forming the duplex with energy  $E_{\text{stackedbases\_dsr}}$  can be regarded as the sum of the Watson–Crick hydrogen-bond interactions ( $\Delta E_{\text{H-bond}}$ )



**Figure 5.** Structural decomposition to evaluate the contributions of the hydrogen-bond interactions  $\Delta E_{\text{H-bond}}$  (dotted lines) and cross terms  $\Delta E_{\text{cross}}$  (arrows) to the total interaction energy  $\Delta E_{\text{int}}$  between two interacting stacks of nucleobases without backbone. Color code in ball-and-stick structures: H – white; C – grey; N – blue; O – red.

and the diagonal interactions or so-called “cross terms” ( $\Delta E_{\text{cross}}$ ), which are the interactions between bases that are diagonal to each other (i.e., on opposite sides and in different layers).

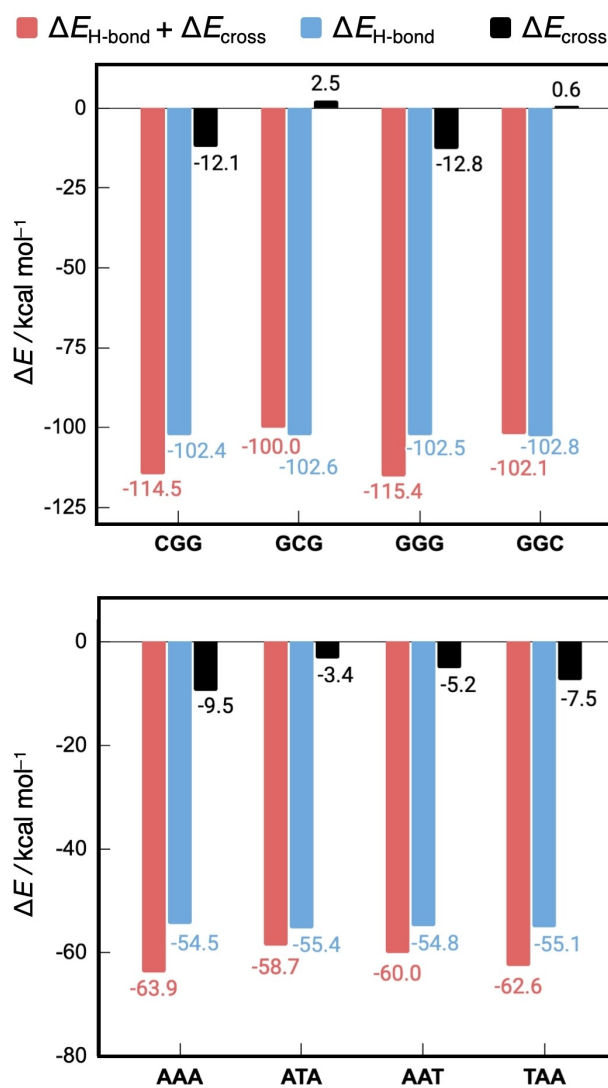
$$\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}} = E_{\text{stackedbases\_ds}} - E_{\text{stackedbases\_ss1'}} - E_{\text{stackedbases\_ss2'}} \quad (12)$$

$\Delta E_{\text{H-bond}}$  is computed as the sum of the energies of the three base pairs ( $E_{\text{bp}}$ , see Figure 5) minus the sum of the energies of the six separate nucleobases in the geometry of the duplex (Equation (13)).  $\Delta E_{\text{cross}}$  is obtained by subtracting  $\Delta E_{\text{H-bond}}$  from the total interaction energy between the two stacks (i.e.,  $\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}}$ ).

$$\Delta E_{\text{H-bond}} = E_{\text{bp1}} + E_{\text{bp2}} + E_{\text{bp3}} - \sum_{i=1}^3 (E_{\text{base\_ss1}'} + E_{\text{base\_ss2}})_i \quad (13)$$

The computed  $\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}}$  and relative contributions are presented in Figure 6 for the all-GC and all-AT sequences. Note that  $\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}}$  in Figure 6 for the DNA systems without the backbone shows the same trend but has not the same magnitude as  $\Delta E_{\text{int}}$  of the complete DNA systems with backbone (Table 2) because the backbone of one strand also interacts weakly with the stacked DNA bases of the other strand. Our computations show that for  $d(\text{CGG})^{\text{H}+}$ , the interaction between the backbone of the CCG strand and the stacked bases of the complementary CGG strand is approximately  $-3 \text{ kcal mol}^{-1}$ . For the  $(\text{CCG})^{\text{H}+}$  single strand with backbone interacting with the complementary CGG single strand without the backbone, a  $\Delta E_{\text{int}}$  of  $-117.1 \text{ kcal mol}^{-1}$  was reported. In this case, only one backbone interacts with the stacked DNA bases of the complementary strand, so that the value of  $-117.1 \text{ kcal mol}^{-1}$  lies in between the interaction energy of the complete duplex  $d(\text{CGG})^{\text{H}+}$  ( $-118.8 \text{ kcal mol}^{-1}$ ) and the CGG duplex without backbone ( $-114.5 \text{ kcal mol}^{-1}$ ). This confirms that the backbone of one strand interacts weakly with the bases of the complementary strand. The significant size of the backbone, and its bent orientation towards the nucleobases, makes this attractive backbone–complementary bases interaction of such significant size that it overcomes the slightly weakening effect of introducing alkyl substituents on the hydrogen bonding interaction in G–C and A–U/T base pairs.<sup>[14a,15]</sup>

From Figure 6 follows that the sequence-dependent variation in the total interaction energy ( $\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}}$ ) originates from the cross-term interactions ( $\Delta E_{\text{cross}}$ ), as the hydrogen-bond energy ( $\Delta E_{\text{H-bond}}$ ) varies barely for the different sequences. Among the GC sequences, the CGG and GGG duplexes experience a more stabilizing  $\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}}$  than GCG and GGC by approximately  $-15 \text{ kcal mol}^{-1}$ . This is in line with the results in Table 2 for the DNA systems with backbone. The more stabilizing total interactions for CGG and GGG follow directly from the very stabilizing cross terms ( $> -12 \text{ kcal mol}^{-1}$ ) for these sequences, while GGC and GCG experience small destabilizing diagonal interactions. The hydrogen-bond interaction is almost constant for all GC sequences.



**Figure 6.** Relative contributions of the hydrogen bonds and cross terms to the total interaction energy ( $\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}}$ ) between two complementary single strands without backbone in the DNA duplex for the all-GC (top) and all-AT (bottom) sequences. Energies are in  $\text{kcal mol}^{-1}$  and computed at the BLYP-D3(BJ)/TZ2P level of theory.

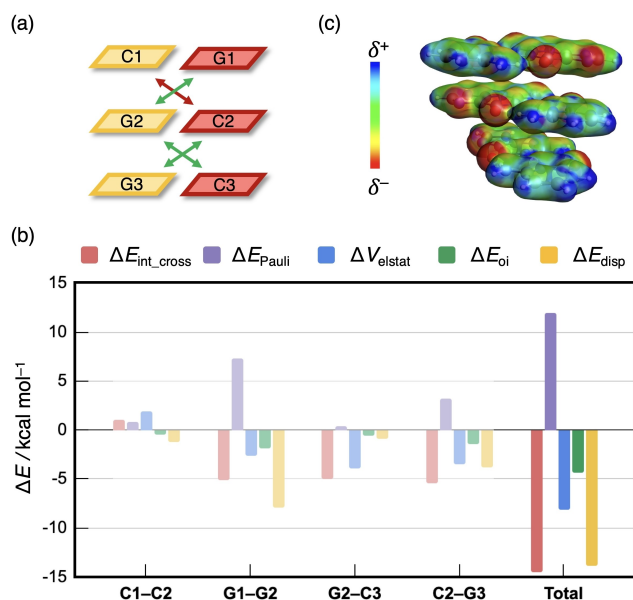
For the AT sequences, the total interaction strength also correlates with the differences in the diagonal interactions:  $\Delta E_{\text{cross}}$  is stabilizing in all cases and the magnitude decreases going from AAA to TAA to AAT to ATA, explaining the trend of  $\Delta E_{\text{H-bond}} + \Delta E_{\text{cross}}$  as  $\Delta E_{\text{H-bond}}$  is nearly constant.

Whether the total cross terms are stabilizing or destabilizing depends on the net electrostatic attraction or repulsion between polar functional groups on diagonal nucleobases. We demonstrate this nucleotide order dependence of  $\Delta E_{\text{cross}}$  by decomposing the diagonal interactions present in the interacting DNA stacks into physically meaningful terms based on Kohn–Sham molecular orbital theory using a quantitative energy decomposition analysis (EDA).<sup>[7]</sup> This analysis decomposes the cross-term interaction energy ( $\Delta E_{\text{int\_cross}}$ ) into Pauli repulsion ( $\Delta E_{\text{Pauli}}$ ), electrostatic interaction ( $\Delta V_{\text{elstat}}$ ), orbital

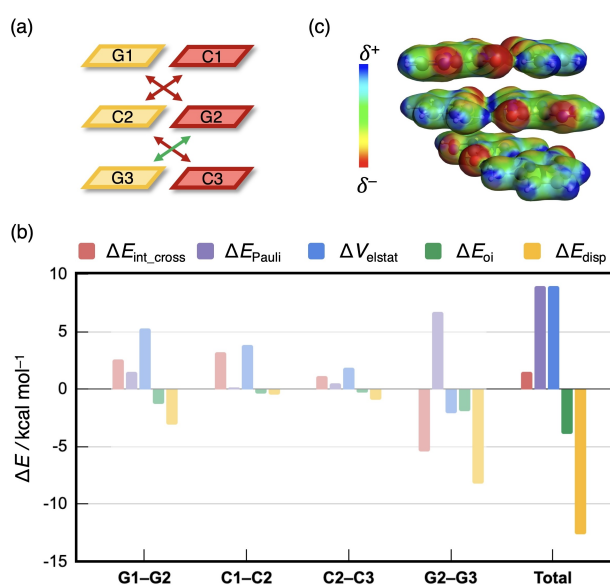
interaction ( $\Delta E_{oi}$ ), and dispersion ( $\Delta E_{disp}$ ) components (Equation (14)).

$$\Delta E_{int\_cross} = \Delta E_{Pauli} + \Delta V_{elstat} + \Delta E_{oi} + \Delta E_{disp} \quad (14)$$

Here,  $\Delta E_{Pauli}$  corresponds to the destabilizing interactions between the occupied orbitals on the fragments and accounts for any steric repulsion.  $\Delta V_{elstat}$  accounts for the classical



**Figure 7.** (a) Three stabilizing and one destabilizing cross-terms are present in  $d(CG)H^+$  without backbone. (b) Partitioning of the cross-term interaction energy  $\Delta E_{int\_cross}$  (in kcal mol<sup>-1</sup>) between diagonal bases in the CGG duplex without backbone, computed at the BLYP-D3(BJ)/TZ2P level. (c) Electrostatic potential surface (at 0.03 au) from  $-0.0$  (red) to  $+0.2$  (blue).



**Figure 8.** (a) One stabilizing and three destabilizing cross-terms are present in  $d(GC)H^+$  without backbone. (b) Partitioning of the cross-term interaction energy  $\Delta E_{int\_cross}$  (in kcal mol<sup>-1</sup>) between diagonal bases in the GCG duplex without backbone, computed at the BLYP-D3(BJ)/TZ2P level. (c) Electrostatic potential surface (at 0.03 au) from  $-0.0$  (red) to  $+0.2$  (blue).

electrostatic interactions between the unperturbed charge distributions of the prepared (i.e., deformed) interacting molecular fragments and is usually attractive. The term  $\Delta E_{oi}$  comprises charge transfer (i.e., donor–acceptor interactions between occupied orbitals on one of the interacting fragments and unoccupied orbitals on the other, including HOMO–LUMO interactions) and polarization (i.e., empty–occupied orbital mixing on one fragment due to the presence of the other fragment). The term  $\Delta E_{disp}$  includes a dispersion energy correction.

To trace the origin of the sequence dependence of the cross terms and thus the total interaction energy between two complementary DNA strands, we focus on the two extreme cases in Figures 7 and 8, that is the CGG and the GCG sequence with stabilizing and destabilizing total cross terms, respectively. For the numerical values of the EDAs, see Supporting Information Figure S6, which also reports the interaction between the bases that are diagonal to each other but not in adjacent layers. This interaction was found to be negligibly small ( $<1$  kcal mol<sup>-1</sup>) and is in line with the popular “nearest neighbor” approach to predict thermodynamic properties of DNA sequences.<sup>[16]</sup>

The decomposition of the diagonal interaction energy within the CGG duplex in Figure 7 shows that this sequence encounters three stabilizing (G1–G2, G2–C3, and C2–G3) and one destabilizing cross terms (C1–C2), and therefore a net stabilizing  $\Delta E_{int\_cross}$  (Figure 7a). The total  $\Delta E_{int\_cross}$  consists, in general, of a large stabilizing dispersion contribution  $\Delta E_{disp}$ , a significant amount of electrostatic interaction  $\Delta V_{elstat}$ , and a smaller contribution of orbital interactions  $\Delta E_{oi}$  counteracted by a large destabilizing Pauli repulsion  $\Delta E_{Pauli}$  (Figure 7b). Whether an individual diagonal interaction is stabilizing or destabilizing is dictated by the electrostatic interaction term. For example, the diagonal interaction between the bases C1 and C2 is destabilizing which results mainly from the destabilizing nature of  $\Delta V_{elstat}$ . This repulsive electrostatic interaction can be visualized by the electrostatic potential surface in Figure 7c. The hydrogen-bond front atoms of the nucleobases guanine and cytosine are  $\delta^+$  NH<sub>2</sub> groups or  $\delta^-$  carbonyl groups. Naturally, two diagonal  $\delta^+$  or two  $\delta^-$  functional groups will slightly repel each other, while opposite charges attract. Following this rationale, the C1–C2 interaction is destabilizing due to the electrostatic repulsion between two  $\delta^+$  NH<sub>2</sub> groups. On the other hand, the stabilizing G2–C3 interaction results from the attractive interaction between a  $\delta^-$  carbonyl and  $\delta^+$  NH<sub>2</sub> group.

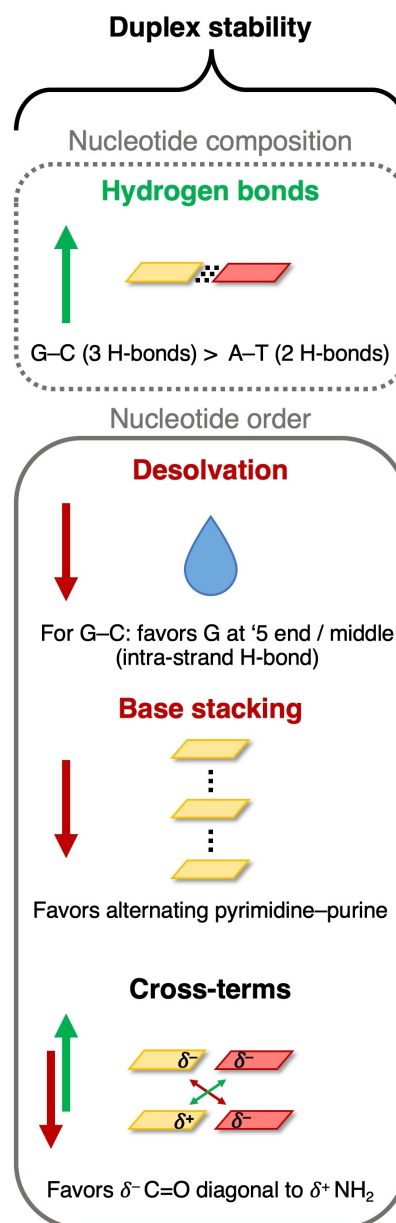
The diagonal interaction energy was also decomposed for the GCG duplex without backbone (Figure 8). This DNA double-strand gives rise to three destabilizing (G1–G2, C1–C2, and C2–C3), and only one stabilizing (G2–G3) cross-term interaction (Figure 8a). This follows from the EDA in Figure 8b, which shows that the diagonal interactions between G1–G2, C1–C2, and C2–C3 are destabilizing due to a significant destabilizing  $\Delta V_{elstat}$ . The electrostatic potential plot in Figure 8c reveals that this repulsive  $\Delta V_{elstat}$  is the result of the unfavorable interaction between two diagonal  $\delta^-$  carbonyl groups, in the case of G1–G2, or between two diagonal  $\delta^+$  amino groups, in the case

of C1–C2 and C2–C3. As the absolute contributions of  $\Delta E_{\text{Pauli}}$ ,  $\Delta E_{\text{oir}}$  and  $\Delta E_{\text{disp}}$  to the total  $\Delta E_{\text{int\_cross}}$  of GCG (Figure 8b) are nearly equal to the contributions for CGG (Figure 7b), we can conclude that  $\Delta V_{\text{elstat}}$  determines whether the total cross-term interaction energy is stabilizing or destabilizing.

The cross-term interaction energy was also decomposed for two all-AT d(DNA)<sup>H+</sup> sequences without backbone, and the results of this analysis can be found in Supporting Information Figure S7. Here, the difference in the total  $\Delta E_{\text{int\_cross}}$  is also dictated by the difference in the electrostatic interaction  $\Delta V_{\text{elstat}}$  resulting from combinations of attractive and repulsive interactions between polar functional groups involved in the Watson–Crick hydrogen bonding.

One of the main objectives of this work is to derive general rules for predicting duplex stability. In this last paragraph, summarized in Figure 9, we translate all findings and analyses into a set of rules to quantitatively predict stabilization or destabilization of DNA duplexes. The terms contributing to the duplex stability are counteracting and vary as a function of the nucleotide composition and order (see Figure 9). The subtle optimum of these terms determines which nucleotide sequence forms the most stable duplex. The hydrogen bonding between complementary bases is the most important stabilizing interaction within the DNA duplex and depends on the nucleotide composition as the triply hydrogen-bonded G–C base pair is more stable than the only doubly hydrogen-bonded A–T base pair. However, as we have shown, the hydrogen-bond interactions are invariant with respect to the nucleotide order, and the sequence-dependent variations in the duplex stability arise from the three other effects, and primarily the cross terms.

First of all, upon forming the duplex, the nucleobases need to be partially desolvated which works destabilizing. As the desolvation energy depends on the type of nucleobase, this term gives rise to sequence-dependent solvent effects. This is particularly relevant for duplexes containing G–C base pairs, as G can form intra-strand hydrogen bonds with the sugar-phosphate backbone when at the 5' end position or somewhere in the middle of the strand, which reduces the required desolvation energy. Furthermore, we identified the loss of stacking interactions within individual strands which is necessary to facilitate the Watson–Crick hydrogen bonding, as a destabilizing factor in the duplex formation. As pyrimidine and purine bases stack less well than bases of the same type, this term prefers an alternating sequence of pyrimidine and purine bases in the DNA double strand. Finally, we have shown that a strong dependence of the duplex stability on the nucleotide order is caused by the cross-term interactions, which can be stabilizing or destabilizing depending on the net electrostatic attraction or repulsion between the hydrogen bonding functional groups on diagonal nucleobases. The positioning of a  $\delta^+$  NH<sub>2</sub> group of one nucleobase opposite to a  $\delta^-$  carbonyl group of its diagonal nucleobase leads to attractive cross terms and, thus, in general, to more stable duplexes.



**Figure 9.** Schematic summary of the terms that contribute to the nucleotide composition and order trends in the stability of duplex DNA assembling from two complementary single strands, with stabilizing terms in green and destabilizing terms in red.

## Conclusions

The stability of double-stranded DNA is not only proportional to the GC base pair content (i.e., more stable than the AT base pair due to an additional hydrogen bond), but also depends on the order of the Watson–Crick base pairs for a given GC versus AT content. This emerges from our quantum chemical analyses using dispersion-corrected DFT computations of all 32 possible double-stranded B-DNA triplets in aqueous implicit solvation, in which we pinpoint the origin of these effects and provide guidelines for predicting the stability of oligonucleotides.

All possible duplex DNA triplets are stable in aqueous solution, also including an estimate of the destabilizing, but roughly constant, thermodynamic corrections (mainly from the zero-point vibrational energy and loss of entropy). Therefore, the trend in duplex stability primarily depends on the electronic energy of formation,  $\Delta E$ . The coordination of counterions to the negatively charged sugar-phosphate backbone stabilizes the overall duplex but was found to not affect the nucleotide-dependent trends significantly. Our activation strain analyses reveal that solvent effects are destabilizing, as duplex formation from complementary single strands requires partial desolvation of the nucleobases.

Although base stacking is often regarded as a crucial contributor to the double helix stability, we here identify the loss of stacking interactions within individual strands, to facilitate the Watson–Crick hydrogen bonding, as a destabilizing factor in the duplex formation. On the other hand, the deformation of the backbone was found to be negligible. The dependence of  $\Delta E$  on the order of the nucleotides for duplexes of the same overall base pair composition follows primarily from the “diagonal interactions” or “cross terms”. We establish that, in general, more stabilizing cross terms lead to more stabilizing interaction energies and, in turn, more stable duplexes. The positioning of a  $\delta^+$   $\text{NH}_2$  group of one nucleobase opposite to a  $\delta^-$  carbonyl group of its diagonal nucleobase manifests in attractive cross terms, while the interaction between two diagonal  $\delta^+$   $\text{NH}_2$  or  $\delta^-$  carbonyl groups is repulsive. These insights have the potential to provide a more unified understanding of genome stability and also provide design principles for oligonucleotides with tailored stabilities.

## Computational Details

All calculations were performed using the Amsterdam Density Functional (version ADF2017.103) program developed by Baerends, Ziegler, and others,<sup>[17]</sup> and the Quantum-regions Interconnected by Local Descriptions (QUILD) program by Swart and Bickelhaupt.<sup>[18]</sup> The QUILD program is a wrapper around ADF (and other programs) and is used for its superior geometry optimizer which is based on adapted delocalized coordinates.<sup>[18a]</sup> The numerical integration was performed using the procedure developed by te Velde et al.<sup>[17a,b]</sup>

All computations are dispersion-corrected density functional theory (DFT-D3) based using the BLYP-D3(BJ) functional in which the regular BLYP functional<sup>[19,20]</sup> is augmented with an empirical correction for long-range dispersion effects, described by a sum of damped interatomic potentials of the form  $C_6R^{-6}$  added to the usual DFT energy.<sup>[21]</sup> The basis set superposition error (BSSE) on the bond energy is effectively absorbed into the empirical dispersion-correction potential.<sup>[21a]</sup>

Molecular orbitals were expanded in a large uncontracted set of Slater-type orbitals (STOs) containing diffuse functions using a TZ2P basis set.<sup>[17c]</sup> The basis set is of triple- $\zeta$  quality for all atoms and has been augmented with two sets of polarization functions, that is,  $2p$ ,  $3d$  on H, and  $3d$  and  $4f$  on C, N, O, Na, and P. The  $1s$  core shells of the atoms C, N, O, Na, and P were treated by the frozen-core approximation. The quality of the density fitting (ZlmFit) and the integration grid (BeckeGrid) was set to “good” for all computations.<sup>[22]</sup> The reported Gibbs free energies in solution are calculated by adding thermal corrections computed at 298 K from

vibrational frequencies obtained through numerical differentiation of the analytical gradient at COSMO( $\text{H}_2\text{O}$ )-ZORA-BLYP-D3(BJ)/TZ2P to the total electronic energy computed at the same level. A more complete description of entropy effects can be obtained through MD simulations of model DNA systems in a box of explicit water molecules. This is, however, beyond the scope of this work in which we aim at estimating the existence of a constant entropic contribution to  $\Delta G$  that does not affect the trends set by  $\Delta E$ .

Solvation in water was simulated using the conductor-like screening model (COSMO), as implemented in the ADF program<sup>[11]</sup> with the solvent radius and dielectric constant being 1.9 Å and 78.4, respectively for water. Atomic radii to generate the COSMO cavities were set equal to the MM3 van der Waals radii divided by a factor 1.2 (for details, see Table S1 in the Supporting Information of Ref. [23]). The surface charges at the GEPOL93 solvent-excluding surface were corrected for outlying charges. According to the work by Riley et al.,<sup>[24]</sup> the dispersion correction does not need to be modified for the solvated systems.

The starting structures of the single and double-stranded B-DNA structures were build using the NUCLEIC routine of the TINKER molecular design program package.<sup>[25]</sup> The analysis of the backbone torsion angles and the Watson–Crick hydrogen bonds in the double-stranded structures was performed using the 3DNA software.<sup>[26]</sup> The optimized structures were illustrated with CYLview20.<sup>[27]</sup>

## Acknowledgements

The work has been performed under the Project HPC-EUROPA3 (INFRAIA-2016-1-730897). The authors gratefully acknowledge the financial support of the EC Research Innovation Action in the Horizon 2020 program and the Netherlands Organization for Scientific Research (NWO). This work was carried out on the Dutch national e-infrastructure with the support of the SURF Cooperative.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

**Keywords:** activation strain model · density functional calculations · diagonal interactions · DNA structures · nucleotide composition

[1] a) J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*, 5th Edition, W. H. Freeman and Company, New York, 2002; b) W. Saenger, *Principles of Nucleic Acid Structure*, Springer, New York, 1984, p. 122.

[2] a) V. Brazda, M. Fojta, R. P. Bowater, *Biochem. J.* 2020, 477, 325–339; b) H. Zeng, M. Mondal, R. Song, J. Zhang, B. Xia, M. Liu, C. Zhu, B. He, Y. Q. Gao, C. Yi, *Angew. Chem. Int. Ed.* 2019, 58, 130–133; *Angew. Chem.* 2019, 131, 136–139; c) L. S. Stelzl, N. Erlenbach, M. Heinz, T. F. Prisner, G.

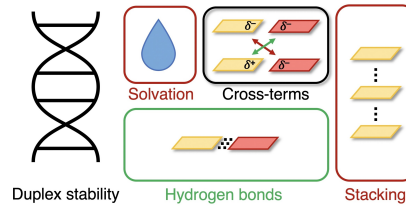
- Hummer, *J. Am. Chem. Soc.* **2017**, *139*, 11674–11677; d) S. Takahashi, N. Sugimoto, *Angew. Chem. Int. Ed.* **2013**, *52*, 13774–13778; *Angew. Chem.* **2013**, *125*, 14019–14023.
- [3] For a recent review, see: S. Takahashi, N. Sugimoto, *Chem. Soc. Rev.* **2020**, *49*, 8439–8468, and references therein.
- [4] a) J. Poater, M. Swart, C. Fonseca Guerra, F. M. Bickelhaupt, *Comput. Theor. Chem.* **2012**, *998*, 57–63; b) D. Soler-Polo, J. I. Mendieta-Moreno, D. G. Trabada, J. Mendieta, J. Ortega, *J. Chem. Theory Comput.* **2019**, *15*, 6984–6991; c) T. A. Hamlin, J. Poater, C. Fonseca Guerra, F. M. Bickelhaupt, *Phys. Chem. Chem. Phys.* **2017**, *19*, 16969–16978; d) J. Poater, M. Swart, F. M. Bickelhaupt, C. Fonseca Guerra, *Org. Biomol. Chem.* **2014**, *12*, 4691–4700; e) V. I. Poltev, V. M. Anisimov, V. I. Danilov, D. Garcia, A. Deriabina, E. González, R. Salazar, F. Rivas, N. Polteva, *Comput. Theor. Chem.* **2011**, *975*, 69–75; f) G. Barone, C. Fonseca Guerra, F. M. Bickelhaupt, *ChemistryOpen* **2013**, *2*, 186–193.
- [5] a) L. Gorb, T. A. Zubatiuk, R. Zubatyuk, D. M. Hovorun, J. Leszczynski, *J. Mol. Model.* **2017**, *23*, 289–301; b) L. Gorb, A. Pekh, A. Nyporko, M. Ilchenko, A. Golius, T. A. Zubatiuk, R. Zubatyuk, I. Dubey, D. M. Hovorun, J. Leszczynski, *J. Phys. Chem. B* **2020**, *124*, 9343–9353; c) Y. Liu, X. Ren, L. He, *J. Chem. Phys.* **2019**, *151*, 215102; d) M. Porrini, F. Rosu, C. Rabin, L. Darré, H. Gómez, M. Orozco, V. Gabelica, *ACS Cent. Sci.* **2017**, *3*, 454–461; e) H. Kruse, M. Havrila, J. Šponer, *J. Chem. Theory Comput.* **2014**, *10*, 2615–2629; f) T. A. Zubatiuk, O. V. Shishkin, L. Gorb, D. M. Hovorun, J. Leszczynski, *Phys. Chem. Chem. Phys.* **2013**, *15*, 18155–18166; g) F. L. Gervasio, P. Carloni, M. Parrinello, *Phys. Rev. Lett.* **2002**, *89*, 108102.
- [6] For reviews, see: a) I. Fernández, F. M. Bickelhaupt, *Chem. Soc. Rev.* **2014**, *43*, 4953–4967; b) F. M. Bickelhaupt, K. N. Houk, *Angew. Chem. Int. Ed.* **2017**, *56*, 10070–10086; *Angew. Chem.* **2017**, *129*, 10204–10221; c) I. Fernández, *Chem. Sci.* **2020**, *11*, 3769–3779. For a step-by-step protocol, see: d) P. Vermeeren, S. C. C. van der Lubbe, C. Fonseca Guerra, F. M. Bickelhaupt, T. A. Hamlin, *Nat. Protoc.* **2020**, *15*, 649–667.
- [7] a) T. A. Hamlin, P. Vermeeren, C. Fonseca Guerra, F. M. Bickelhaupt, in *Complementary Bonding Analyses* (Eds.: S. Grabowsky), De Gruyter, Berlin, **2021**, pp. 199–212; b) F. M. Bickelhaupt, E. J. Baerends, in *Reviews in Computational Chemistry*, Vol. 15 (Eds.: K. B. Lipkowitz, D. B. Boyd), Wiley-VCH, New York, **2000**, pp. 1–86; c) L. Zhao, M. von Hopffgarten, D. M. Andrada, G. Frenking, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1345.
- [8] a) C. Nieuwland, F. Zaccaria, C. Fonseca Guerra, *Phys. Chem. Chem. Phys.* **2020**, *22*, 21108–21118; b) C. Fonseca Guerra, T. van der Wijst, J. Poater, M. Swart, F. M. Bickelhaupt, *Theor. Chem. Acc.* **2010**, *125*, 245–252; c) F. Zaccaria, S. C. C. van der Lubbe, C. Nieuwland, T. A. Hamlin, C. Fonseca Guerra, *ChemPhysChem* **2021**, *22*, 2286–2296; d) J. Šponer, A. Mládek, N. Špačková, X. Cang, T. E. Cheatham III, S. Grimme, *J. Am. Chem. Soc.* **2013**, *135*, 9785–9796; e) T. van der Wijst, C. Fonseca Guerra, M. Swart, F. M. Bickelhaupt, B. Lippert, *Angew. Chem. Int. Ed.* **2009**, *48*, 3285–3287; *Angew. Chem.* **2009**, *121*, 3335–3337; f) L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.
- [9] a) R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, P. Hobza, *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374; b) N. Mardirossiana, M. Head-Gordon, *Mol. Phys.* **2017**, *115*, 2315–2372.
- [10] B. Schneider, S. Neidle, H. M. Berman, *Biopolymers* **1997**, *42*, 113–124.
- [11] a) A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2* **1993**, *2*, 799–805; b) A. Klamt, *J. Phys. Chem.* **1995**, *99*, 2224–2235; c) C. C. Pye, T. Ziegler, *Theor. Chem. Acc.* **1999**, *101*, 396–408.
- [12] a) A. Vologodskii, M. D. Frank-Kamenetskii, *Phys. Life Rev.* **2018**, *25*, 1–21; b) J. Černý, M. Kabeláč, P. Hobza, *J. Am. Chem. Soc.* **2008**, *130*, 16055–16059.
- [13] M. Zacharias, *J. Phys. Chem. B* **2020**, *124*, 10345–10352.
- [14] a) C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, *J. Am. Chem. Soc.* **2000**, *122*, 4117–4128; b) C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, *Chem. Eur. J.* **1999**, *5*, 3581–3594; c) C. Fonseca Guerra, F. M. Bickelhaupt, *Angew. Chem. Int. Ed.* **1999**, *38*, 2942–2945; *Angew. Chem.* **1999**, *111*, 3120–3122.
- [15] M. Swart, C. Fonseca Guerra, F. M. Bickelhaupt, *J. Am. Chem. Soc.* **2004**, *126*, 16718–16719.
- [16] a) G. Gotoh, Y. Tagashira, *Biopolymers* **1981**, *20*, 1033–1042; b) J. M. Huguet, M. Ribezzi-Crivellari, C. V. Bizarro, F. Ritort, *Nucleic Acids Res.* **2017**, *45*, 12921–12931; c) A. Altun, M. Garcia-Ratés, F. Neese, G. Bistoni, *Chem. Sci.* **2021**, *12*, 12785–12793.
- [17] a) G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931–967; b) E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41–51; c) E. van Lenthe, E. J. Baerends, *J. Comput. Chem.* **2003**, *24*, 1142–1156; d) C. Fonseca Guerra, J. G. Snijders, G. te Velde, E. J. Baerends, *Theor. Chem. Acc.* **1998**, *99*, 391–403; e) E. J. Baerends et al., ADF2017.103, SCM, Theoretical Chemistry, Vrije Universiteit Amsterdam, Amsterdam (The Netherlands), **2017** www.scm.com.
- [18] a) M. Swart, F. M. Bickelhaupt, *Int. J. Quantum Chem.* **2006**, *106*, 2536–2544; b) M. Swart, F. M. Bickelhaupt, *J. Comput. Chem.* **2008**, *29*, 724–734.
- [19] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [20] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
- [21] a) S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463–1473; b) S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787–1799; c) J. Antony, S. Grimme, *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293; d) M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, *J. Chem. Phys.* **2001**, *114*, 5149–5155; e) X. Wu, M. C. Vargas, S. Nayak, V. Lotrich, G. Scoles, *J. Chem. Phys.* **2001**, *115*, 8748–8757; f) Q. Wu, W. Yang, *J. Chem. Phys.* **2001**, *116*, 515–524; g) U. Zimmerli, M. Parrinello, P. Koumoutsakos, *J. Chem. Phys.* **2004**, *120*, 2693–2699.
- [22] a) M. Franchini, P. H. T. Philipsen, E. van Lenthe, L. Visscher, *J. Chem. Theory Comput.* **2014**, *10*, 1994–2004; b) M. Franchini, P. H. T. Philipsen, L. Visscher, *J. Comput. Chem.* **2013**, *34*, 1819–1827.
- [23] M. Swart, E. Rösler, F. M. Bickelhaupt, *Eur. J. Inorg. Chem.* **2007**, 3646–3654.
- [24] K. E. Riley, J. Vondrášek, P. Hobza, *Phys. Chem. Chem. Phys.* **2007**, *9*, 5555–5560.
- [25] J. W. Ponder, TINKER, Software Tools for Molecular Design, Washington University, Saint Louis, Missouri (USA), **2017**, <http://dasher.wustl.edu/tinker/>.
- [26] a) X. J. Lu, W. K. Olson, *Nat. Protoc.* **2008**, *3*, 1213–1227; b) X. J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121; c) X. J. Lu, W. K. Olson, 3DNA, Rutgers University, Piscataway, New Jersey (USA), <https://x3dna.org/>.
- [27] C. Y. Legault, CYLview20, Visualization and analysis software for computational chemistry, Université de Sherbrooke, Sherbrooke, Quebec (Canada), **2020**, www.cylview.org.

Manuscript received: October 12, 2021

Revised manuscript received: December 10, 2021

# RESEARCH ARTICLE

**All about that sequence!** Quantum chemical analyses reveal that the stability of B-DNA triplets depends not only on the nucleotide composition, but also on the order in which the nucleotides occur. This sequence dependence in the duplex stability results primarily from attractive or repulsive cross terms between bases in adjacent Watson–Crick pairs.



*C. Nieuwland, Dr. T. A. Hamlin,  
Prof. Dr. C. Fonseca Guerra\*,  
Prof. Dr. G. Barone\*, Prof. Dr. F. M.  
Bickelhaupt\**

1 – 15

**B-DNA Structure and Stability: The  
Role of Nucleotide Composition and  
Order**

