The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/123184

Please be advised that this information was generated on 2020-02-20 and may be subject to change.
B-DNA Structure and Stability as Function of Nucleic Acid Composition: Dispersion-Corrected DFT Study of Dinucleoside Monophosphate Single and Double Strands

Giampaolo Barone,*[a] Célia Fonseca Guerra,[b] and F. Matthias Bickelhaupt*[b, c]

We have computationally investigated the structure and stability of all 16 combinations of two out of the four natural DNA bases, A, T, G, and C, in a di-2'-deoxyribonucleoside-monophosphate DNA strand as well as in 10 double-strand model complexes thereof, using dispersion-corrected density functional theory (DFT-D). Optimized geometries with B-DNA conformation were obtained through the inclusion of implicit water solvent and, in the DNA models, of sodium counterions, to neutralize the negative charge of the phosphate groups. The results obtained allowed us to compare the relative stability of isomeric single and double strands. Moreover, the energy of the Watson–Crick pairing of complementary single strands to form double-helical structures was calculated. The latter furnished the following increasing stability trend of the double-helix formation energy: \(d(TpA) < d(CpA) < d(ApT) < d(ApA)\), \(d(GpT) < d(GpA) < d(ApG) < d(CpG) < d(GpG)\), where the energy differences between the last four dimers, \(d(ApG)_v\), \(d(CpG)_v\), \(d(GpG)_v\), and \(d(GpC)_v\), is within 4.0 kcal mol\(^{-1}\). and the energy between the most and the least stable isomers is 13.4 kcal mol\(^{-1}\). This trend shows that the formation energy essentially increases with the number of hydrogen bonds per base pair, that is, two between A and T and three between G and C. Superimposed on this main trend are more subtle effects that depend on the order in which bases occur within a strand from the 5’- to the 3’-end.

Introduction

The DNA double helix results from several types of stabilizing but also destabilizing interactions between and within its two complementary, intertwined polydeoxyribonucleotide chains. In fact, the major contributions to the stability of the DNA structure come from:[21] (1) hydrogen bonds in adenine–thymine (AT) and guanine–cytosine (GC) Watson–Crick pairs; (2) \(\pi-\pi\) stacking interactions among Watson–Crick paired nitrogen bases stacked along the double helix axis; (3) electrostatic repulsive interactions among the phosphate groups; (4) electrostatic attractive interactions between the phosphate groups and cations dissolved in water solution; (5) hydrophilic interactions of the sugar-phosphate skeleton with water; (6) hydrophobic interactions of the DNA cylindrical core, made up by the hydrogen-bonded and stacked nitrogen bases, with the water solvent. Recently, there has been increasing effort in developing and applying quantum chemical methods able to reproduce the structure of native B-DNA and to correctly describe the energy involved in the intrastrand and interstrand noncovalent interactions between the nucleotide monomers. This topic has been approached by both wave function methods and density functional theory.[25]

Water solvent and sodium counterions also play an important role in the formation and relative stabilization of the double-helical DNA structure because they mitigate the aforementioned long-range electrostatic repulsions among the phosphate groups in the DNA backbone and amplify \(\pi-\pi\) stacking interactions through the hydrophobic effect. This issue has been recently considered in the choice of model systems in density functional theory (DFT) calculations of the structure of dinucleoside monophosphate single strands.[23] The results obtained show that, not unexpectedly, the presence of the Na\(^+\) counterions at each phosphate group is even more important than the presence of the implicit solvent for providing a structure in which the relative base–base orientation and the backbone torsion angles are in better agreement with experimental B-DNA crystal structures. This result reconfirms the importance of charge neutralization in DNA model systems.

A complicating aspect of DNA in computational studies is the combination of the large size of model systems in combination with the high demand on accuracy for describing the
various delicate, yet crucial types of interactions, such as, hy-
drogen bonding and π–π stacking. The latter can be adequate-
ly described using DFT or post-Hartree–Fock ab initio methods
that properly account for electron correlation. However, the
size already of minimal model systems for double-stranded
DNA is prohibitive for post-Hartree–Fock approaches. For ex-
ample, coupled-cluster theory with single and double electron
excitations and triple electron excitations treated perturbative-
ly, that is, CCSD(T), at the estimated complete basis set (CBS)
limit, can be applied to molecules containing no more than
about 50 atoms. On the other hand, DFT methods are able
to deal with, and accurately describe these large DNA models.
Hydrogen bonding in Watson–Crick base pairs and other DNA
structures is well described using the generalized gradient ap-
proximation (GGA). Interactions, such as π–π stacking, in
which London dispersion features prominently, require the use
of semilocal hybrid functionals that recover medium-range cor-
relation effects as proposed, for example, by Troullin and cow-
orkers, but they can also be treated very efficiently using
Grimme’s empirically dispersion-corrected density functional
theory (DFT-D).

Model systems in pioneering DFT studies so far cover single-
stranded GC dinucleoside monophosphate models (including
sodium counterions and implicit water solvent), the ten pos-
sible combinations of DNA bases A, T, G and C in stacks of two
Watson–Crick pairs (but without sugar-phosphate bridge be-
tween the Watson–Crick pairs), and double-stranded DNA
dimers (but with the sugar-phosphate bridge in the MM part
of a QM/MM approach). In the present work, we extend the above quest to a quan-
tum chemical exploration of the 16 possible single-stranded
and 10 double-stranded B-DNA models that exist for the dinu-
cleoside monophosphates of A, T, G and C, using dispersion-
corrected DFT at BLYP-D/TZ2P as implemented in the Amster-
dam Density Functional (ADF) program (see below). The latter
approach has been shown to accurately reproduce CCSD(T)
stuctural and energy data of stacked as well as hydrogen-
bonded AT and GC complexes. Our DFT-D study also covers
the sodium counterions at the phosphate groups as well as the
effect of aqueous solvation using the conductor-like screening
model (COSMO). The purpose of our work is to obtain a more
detailed understanding of the structure and sta-
bility of B-DNA and how this depends on DNA-base sequences.
The resulting insight is also relevant, among other things, for
modeling and rationalizing the interaction of drugs with DNA,
following any type of covalent and/or noncovalent drug–
DNA binding with particular emphasis on DNA intercalators.

**Computational Details**

All calculations were performed using the Amsterdam Density
Functional (ADF) program developed by Baerends, Ziegler, and
others, and the QUantum-regions Interconnected by Local
Descriptions (QUILD) program by Swart and Bickelhaupt. 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. The numerical integra-
tion was performed using the procedure developed by te
Velde et al.

Molecular orbitals (MOs) were expanded in a large uncon-
tracted set of Slater-type orbitals (STOs) containing diffuse
functions: TZ2P (no Gaussian functions are involved). The
basis set is of triple-ζ quality for all atoms and has been aug-
mented with two sets of polarization functions, that is, 3d and
4f on C, N, O and 2p, 3d on H. The 1s core shells of carbon, ni-
trogen and oxygen were treated by the frozen-core approxi-
mation. An auxiliary set of s, p, d, f and g STOs was used to fit
the molecular density and to represent the Coulomb and ex-
change potentials accurately in each self-consistent field cycle.

Geometries and energies were computed with dispersion-
corrected density functional theory (DFT-D) using the BLYP-D
functional in which the regular BLYP functional is aug-
mented with an empirical correction for long-range dispersion
effects, described by a sum of damped interatomic potentials
of the form C_R^1/n added to the usual DFT energy. The
basis set superposition error (BSSE) on the bond energy is
effectively absorbed into the empirical dispersion-correction po-
tential. The starting B-DNA structures of the single
and double strands were built using the NUCLEIC routine of the
TINKER molecular design program package.

Solvent effects in water have been estimated using the con-
ductor-like screening model (COSMO), as implemented in the
ADF program with the following settings: solvent radius and
dielectric constant for water were 1.9 Å and 78.4, respectively,
and atomic radii were taken from the MM3 van der Waals radii
scaled by 0.8333 (for details, see Table S1 of ref. [19]). The sur-
face charges at the GEPOL93 solvent-excluding surface were
corrected for outlying charges. According to the work by Riley
et al., the dispersion correction does not need to be modi-
fied for the solvated systems.

The Cartesian coordinates and energies of stationary points
are given in the Supporting Information.

**Results and Discussion**

In the following four sections, we focus on two important find-
ings of our investigations. The first one concerns the reliability
of DFT-D in combination with the conductor-like screening
model (COSMO) for mimicking bulk solvation in aqueous solu-
tion, for the calculation of dinucleoside monophosphate single
and double strands in physiological conditions. The calculated
structures are in fact in good agreement with experimental
structures of B-DNA oligodeoxynucleotides obtained by X-ray
crystallography. The second important finding emerges from
the analyses of the relative stability of the various model B-
DNA double strands in terms of their formation energy from
the corresponding single-stranded moieties as well as the
energy contributions into which the formation energy can be
decomposed. These findings shed light, among others, on how
the overall stability and structure of B-DNA depends on its
composition in terms of AT versus GC Watson–Crick pairs and
the role of intramolecular interactions in genetic evolution.

There are 16 different doublet sequences and this number
increases to 64 for triplet sequences in single-stranded DNA.
Both combinations decrease to 10 and to 32 for doublet and triplet sequences, respectively, in double-helical DNA. In this respect, it is worth recalling that each of the DNA triplet sequences are the smallest units of genetic code in protein synthesis.\cite{22} In the present work, we have focused our attention on DNA dimers, and we plan to extend our investigation on the structure and stability of DNA triplets in a forthcoming paper, using the same computational approach reported herein.

Optimized structures of single- and double-stranded B-DNA dimers

The structures of the 16 possible combinations of the four DNA nitrogen bases A, C, G and T used to build dinucleoside phosphate single strands are reported in Figure 1. The combinations of the four nitrogen bases, giving rise to 10 possible DNA doublet sequences, obtained by Watson–Crick pairing of the complementary single strands above, are reported in Figure 3–5. The geometries of both single and double strands have been fully optimized within the implicit solvent, and their corresponding Cartesian coordinates and absolute energy values are reported in Table S1 in the Supporting Information. The backbone torsion angles of the optimized structures are reported in Tables 1 and 2 and compared with available experimental crystal structures of B-DNA oligonucleotides.\cite{21} The latter derive from a statistical analysis of 34 B-DNA oligodeoxynucleotide structures, characterized with resolutions between 0.7 and 3.3 Å. The hydrogen bond distances between Watson–Crick complementary bases are reported in Table 3. The atom labels used for defining the backbone torsion angles and the hydrogen bonding distances are shown in Figure 2.

There is satisfactory agreement between the average values of the torsion angles of both single and double strands with available experimental crystal structures of oligodeoxynucleotides in B-DNA conformation.\cite{21} The torsion angles and hydrogen bond distances data reported in Tables 1 and 2 show that larger distortions from the B-DNA conformations, with a significantly lower value of the \( \beta \) torsion angle, have been observed for \( d(GpA) \), \( d(GpC) \) and \( d(GpG) \) in single strands and for \( d(GpA) \), \( d(CpA) \) and \( d(CpG) \) in double strands.

The hydrogen bond lengths obtained are also in good agreement with those calculated with the same DFT-D method and basis set for isolated DNA nitrogen bases.\cite{4g} However, the presence of the sugar phosphate chain induces small distortions in the hydrogen-bonded base pairs. These distortions are somewhat more pronounced in the case of AT than GC base pairs. For example, there is a closer match between the distances of isolated AT base pairs and those in \( d(TpA) \), whereas \( N6–O4 \) and \( N1–N3 \) distances are slightly shorter and longer, by 0.1 Å, respectively, in \( d(ApT) \) (see Table 3 and ref. \[4g\]). On the other hand, there is an excellent agreement, within 0.03 Å, between the three \( O6–N4, N1–N3 \) and \( N2–O2 \) distances of isolated GC base-pairs and those in \( d(GpC) \), \( d(CpG) \) and \( d(GpG) \) (see Table 3 and ref. \[4g\]).

Double-strand stability

In the following, we show how the relative stability of the double-stranded dimers emerges from an interplay between the strain energy in the single strands and the interaction between these single strands. Especially trends in the latter can be well understood in terms of molecular structural features. We analyze the stability of a DNA double helix in terms of the (relative) complexation energy associated with forming the double strand from two single strands, \( \Delta E_{\text{double}} \) (or \( \Delta E_{\text{double,rel}} \)), see Equations 1:

\[
\Delta E_{\text{double}} = E[ds] - E[ss1] - E[ss2]
\]

(1a)
This analysis leads to a global order of stabilities among all of our DNA double strands which are collected in Table 3. The major categorization in terms of stability order can be made, not unexpectedly, in terms of the number of AT and GC Watson–Crick pairs. The stability $\Delta E_{\text{double}}$ of the model double helix increases as the number of GC pairs increases from zero (−14 to −19 kcal mol$^{-1}$) to one (−19 to −23 kcal mol$^{-1}$) to two (−24 to −27 kcal mol$^{-1}$). Differences in energy between isomeric single strands are significantly smaller, that is, less than 1 kcal mol$^{-1}$ between TpG and GpT, between ApA and ApC, between GpA and ApG, and between TpA and TpA; 1.2 kcal mol$^{-1}$ between CpT and TpC, and 1.6 kcal mol$^{-1}$ between CpG and GpC (see Table S1).

However, also within a family of isomers, consisting of the same number of AT and GC pairs, there are significant fluctuations in stability. As can be seen in Table 3, the following orders of stability in terms of complexation energy $\Delta E_{\text{double}}$ can be observed for the three families of model double helices (see Equations 2):

\[
d(\text{ApA})_2 > d(\text{ApT})_2 > d(\text{TpA})_2 \quad (2a)
\]
\[
d(\text{ApG})_2 > d(\text{GpA})_2 > d(\text{GpT})_2 > d(\text{ApC})_2 \quad (2b)
\]
\[
d(\text{GpC})_2 > d(\text{GpG})_2 > d(\text{GpT})_2 > d(\text{CpG})_2 \quad (2c)
\]

This is in line with our earlier finding that the affinity of a model template–primer complex for an incoming new DNA base not only depends on the complexation energy of the new Watson–Crick (or mismatched) base pair but also on the terminal base in the primer strand to which the incoming DNA base binds via $\pi$–$\pi$ stacking interactions.$^{[D]}$

### Analysis of double-strand interaction

To obtain a better understanding of the in silico observed trends, we have decomposed the formation energy $\Delta E_{\text{double}}$ of a double strand from two complementary single strands, into two main energy contributions: 1) the strain $\Delta E_{\text{strain}}$ that is the energy required to distort the isolated single strands from their individual equilibrium structure into the geometry they assume in the double strand; and 2) the interaction energy $\Delta E_{\text{int}}$ that is the difference between the energy of the double helix and that of the two single strands with the same geometry they assume in the double helix (see Equation 3):

\[
\Delta E_{\text{double}} = \Delta E_{\text{strain}} + \Delta E_{\text{int}} \quad (3)
\]

The results of our analyses for the 10 double helix models are collected in Table 3. In the first place, it appears that, for a given family of double helical isomers, the trend in relative stability $\Delta E_{\text{double}}$ always follows the amount of strain energy $\Delta E_{\text{strain}}$ associated with deforming the single strands to the geometry they have to adopt to form Watson–Crick pairs in the overall system. For example, along $d(\text{ApA})_2$, $d(\text{ApT})_2$, and $d(\text{TpA})_2$, the strain energy $\Delta E_{\text{strain}}$ increases from 2.9 to 3.6 to 10.3 kcal mol$^{-1}$ and, accordingly, the complexation energy $\Delta E_{\text{double}}$ weakens from −19.1 to −18.7 to −13.8 kcal mol$^{-1}$ (see Table 3). This interesting trend is related to the increasing structural distortion of the complementary single strands when they form the double-stranded complex. In general, all the values of torsion angles of the free single-stranded structures, reported in Table 1, change when they form the double-stranded structures (see Table 2). Remarkably, the value of $\beta$ of the single-stranded doublets, which involves the atoms $C4'–C5'–O5'$ (see Figure 2), experiences the largest modification in the less stable double-stranded structures, namely $d(\text{TpA})_2$, $d(\text{CpA})_2$, and $d(\text{GpG})_2$ (see Table 3). In more detail, the value of $\beta$ changes by some $34^\circ$ for $d(\text{TpA})_2$, $38^\circ$ for $d(\text{TpA})_2$, $48^\circ$ for $d(\text{CpG})_2$, and $52^\circ$ for $d(\text{CpA})_2$ (compare Tables 1 and 2). Inspection of Figure 3–5 (top view) reveals that, in all four of these dinucleotide strands, the associated internal rotations around the $C5'–O5'$ bond leads to a less favorable $\pi$–$\pi$ stacking ar-

---

**Table 1. Backbone torsion angles [\(^\circ\)] of the 16 dinucleoside monophosphate B-DNA single strands**

<table>
<thead>
<tr>
<th>ss</th>
<th>PO3'–C3'–C4' ((\alpha))</th>
<th>OS5'–PO3'–C3' ((\gamma))</th>
<th>C5’–OS5’–PO3' ((\beta))</th>
<th>C4’–C5’–OS5’ ((\gamma))</th>
<th>C3’–C4’–C5’ ((\beta))</th>
<th>C2–N1/C4–N9 ((\gamma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>d(\text{ApT})_2</td>
<td>160.6</td>
<td>−84.5</td>
<td>−61.3</td>
<td>−176.2</td>
<td>51.0</td>
<td>−101.3 (6)</td>
</tr>
<tr>
<td>d(\text{ApA})_2</td>
<td>−177.1</td>
<td>−78.0</td>
<td>−61.0</td>
<td>168.9</td>
<td>50.1</td>
<td>−98.8 (3)</td>
</tr>
<tr>
<td>d(\text{TpT})_2</td>
<td>−174.4</td>
<td>−77.0</td>
<td>−64.6</td>
<td>167.5</td>
<td>50.6</td>
<td>−112.6 (3)</td>
</tr>
<tr>
<td>d(\text{TpA})_2</td>
<td>−175.3</td>
<td>−78.2</td>
<td>−62.8</td>
<td>168.5</td>
<td>46.7</td>
<td>−95.8 (6)</td>
</tr>
<tr>
<td>d(\text{GpT})_2</td>
<td>154.0</td>
<td>−80.7</td>
<td>−60.1</td>
<td>174.1</td>
<td>48.3</td>
<td>−101.6 (6)</td>
</tr>
<tr>
<td>d(\text{ApC})_2</td>
<td>158.1</td>
<td>−83.8</td>
<td>−62.5</td>
<td>−170.2</td>
<td>50.2</td>
<td>−104.9 (8)</td>
</tr>
<tr>
<td>d(\text{ApG})_2</td>
<td>−158.5</td>
<td>−70.4</td>
<td>−64.8</td>
<td>174.7</td>
<td>54.6</td>
<td>−114.0 (32)</td>
</tr>
<tr>
<td>d(\text{CpT})_2</td>
<td>−174.0</td>
<td>−75.0</td>
<td>−65.1</td>
<td>171.5</td>
<td>50.3</td>
<td>−112.0 (1)</td>
</tr>
<tr>
<td>d(\text{GpA})_2</td>
<td>159.7</td>
<td>−85.0</td>
<td>−81.1</td>
<td>−127.7</td>
<td>42.9</td>
<td>−102.5 (15)</td>
</tr>
<tr>
<td>d(\text{TpC})_2</td>
<td>169.4</td>
<td>−90.7</td>
<td>−68.6</td>
<td>161.8</td>
<td>40.0</td>
<td>−109.9 (6)</td>
</tr>
<tr>
<td>d(\text{CpA})_2</td>
<td>−163.9</td>
<td>−68.8</td>
<td>−62.5</td>
<td>175.1</td>
<td>53.9</td>
<td>−102.8 (21)</td>
</tr>
<tr>
<td>d(\text{GpC})_2</td>
<td>−162.1</td>
<td>−72.3</td>
<td>−64.0</td>
<td>175.5</td>
<td>53.1</td>
<td>−106.5 (29)</td>
</tr>
<tr>
<td>d(\text{GpG})_2</td>
<td>166.2</td>
<td>−85.3</td>
<td>−79.5</td>
<td>−134.0</td>
<td>42.2</td>
<td>−116.9 (10)</td>
</tr>
<tr>
<td>d(\text{CpG})_2</td>
<td>−163.0</td>
<td>−71.6</td>
<td>−62.5</td>
<td>175.9</td>
<td>53.3</td>
<td>−104.6 (27)</td>
</tr>
<tr>
<td>d(\text{CpC})_2</td>
<td>160.3</td>
<td>−82.2</td>
<td>−82.3</td>
<td>−123.6</td>
<td>47.9</td>
<td>−100.6 (25)</td>
</tr>
<tr>
<td>d(\text{CpG})_2</td>
<td>168.0</td>
<td>−84.1</td>
<td>−79.4</td>
<td>−141.7</td>
<td>44.7</td>
<td>−111.7 (7)</td>
</tr>
</tbody>
</table>

Av (SD)$^{[a]}$ 176.8 (16) −79.2 (6) −67.6 (8) −170.2 (24) 48.8 (4) −106.1 (17)

Exptl (SD)$^{[b]}$ 184 (11) −95 (10) −62 (15) 176 (9) 48 (11) −102 (14)

[a] Computed at BLYP-D/TZ2P and COSMO for simulating aqueous solution. See Figure 2 for definitions and Figure 1 for structures. [b] Average of the two $\gamma$ angles present in each dinucleoside monophosphate (see Figure 2). [c] Average with standard deviation (SD) over the 16 dinucleoside monophosphate single strands. [d] Experimental values with standard deviation (SD) from X-ray crystal structures of oligodeoxynucleotides in B-DNA conformation (ref. [21]).
rangement between the two stacked DNA bases in the resulting double helices than in the separate single strands. Likewise, the more stable double-stranded isomers with the less strained single-strand fragments display more favorable π-π stacking arrangements between the stacked DNA bases within a single strand. This can be understood as a conflict, arising from the structural characteristics of the dinucleotides d(TpG), d(TpA), d(GpG) and d(CpA), between, on one hand, the requirement to adopt a proper arrangement for entering into inter-strand Watson–Crick pairing and, on the other hand, an optimal intrastrand π-π stacking arrangement. The interaction energy ΔEint behaves in a less systematic manner than ΔEstrain. Along the same series, d(ApA)2, d(ApT)2 and d(TpA)2, ΔEint first remains essentially constant at ~22.1 and ~22.1 kcal mol⁻¹ and then strengthens to ~24.0 kcal mol⁻¹.

To quantify the effect of the backbone on the stability of the Watson–Crick pairs in the duplex structure, we also compare them with those in simple Watson–Crick pairs and in the simple stacks of Watson–Crick pairs, that is, stacks without a sugar–phosphate–sugar backbone. Previously, we found that ΔE_double in aqueous solution for simple AT and GC Watson–Crick pairs amounts to ~9.8 and ~13.6 kcal mol⁻¹, respectively, computed at the same level of theory as used in this study.⁴⁻⁶ If we compare this to the present model duplexes, we see that the latter have effectively slightly weaker Watson–Crick bonds per base pair. For example, in the strongest AT bound duplex, that is, d(ApA)2, the effective Watson–Crick bond strength per base pair amounts to ΔE_double/2 = ~9.6 kcal mol⁻¹. Apparently, the slight increase in deformation strain ΔE_strain that occurs if a backbone is present somewhat reduces the effective stability of the Watson–Crick interaction. As this strain becomes larger, in d(TpA)2, the effective stability is further reduced to ΔE_double/2 = ~6.9 kcal mol⁻¹.

### Table 2. Backbone torsion angles [⁴°] of the 10 dinucleoside monophosphate B-DNA double strands.

<table>
<thead>
<tr>
<th>ds</th>
<th>s [⁴°]</th>
<th>PO3–C3’– C4’ (α)</th>
<th>O5’–PO3– C3’ (β)</th>
<th>C5’–O5’– C4’–C5’ (γ)</th>
<th>C3’–C4’– C5’–O5’ (γ)</th>
<th>C2’–N1/C4’–N9–C3’ (γ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d(ApT)2</td>
<td>d(ApT)</td>
<td>154.5</td>
<td>–81.8</td>
<td>–60.9</td>
<td>–173.3</td>
<td>50.9</td>
</tr>
<tr>
<td>d(ApT)2</td>
<td>d(ApT)</td>
<td>153.6</td>
<td>–81.3</td>
<td>–60.9</td>
<td>–172.6</td>
<td>50.6</td>
</tr>
<tr>
<td>d(ApA)2</td>
<td>d(ApA)</td>
<td>158.0</td>
<td>–82.4</td>
<td>–63.1</td>
<td>–171.8</td>
<td>48.4</td>
</tr>
<tr>
<td>d(TpA)2</td>
<td>d(TpA)</td>
<td>164.8</td>
<td>–86.8</td>
<td>–62.9</td>
<td>–174.2</td>
<td>50.7</td>
</tr>
<tr>
<td>d(TpA)2</td>
<td>d(TpA)</td>
<td>164.1</td>
<td>–90.1</td>
<td>–69.6</td>
<td>–152.7</td>
<td>45.4</td>
</tr>
<tr>
<td>d(TpA)2</td>
<td>d(TpA)</td>
<td>164.9</td>
<td>–90.2</td>
<td>–70.4</td>
<td>–152.5</td>
<td>42.2</td>
</tr>
<tr>
<td>d(GpT)2</td>
<td>d(GpT)</td>
<td>148.4</td>
<td>–79.4</td>
<td>–59.7</td>
<td>–169.2</td>
<td>51.8</td>
</tr>
<tr>
<td>d(GpT)2</td>
<td>d(GpT)</td>
<td>148.9</td>
<td>–79.8</td>
<td>–59.1</td>
<td>–170.4</td>
<td>51.5</td>
</tr>
<tr>
<td>d(ApG)2</td>
<td>d(ApG)</td>
<td>150.9</td>
<td>–79.2</td>
<td>–62.2</td>
<td>–167.7</td>
<td>48.4</td>
</tr>
<tr>
<td>d(ApG)2</td>
<td>d(ApG)</td>
<td>151.6</td>
<td>–87.4</td>
<td>–64.8</td>
<td>–164.6</td>
<td>49.3</td>
</tr>
<tr>
<td>d(GpA)2</td>
<td>d(GpA)</td>
<td>158.2</td>
<td>–84.7</td>
<td>–78.5</td>
<td>–131.5</td>
<td>43.5</td>
</tr>
<tr>
<td>d(CpG)2</td>
<td>d(CpG)</td>
<td>169.9</td>
<td>–92.2</td>
<td>–67.6</td>
<td>–162.0</td>
<td>48.3</td>
</tr>
<tr>
<td>d(CpA)2</td>
<td>d(CpA)</td>
<td>164.0</td>
<td>–85.9</td>
<td>–78.5</td>
<td>–132.8</td>
<td>41.0</td>
</tr>
<tr>
<td>d(TpG)2</td>
<td>d(TpG)</td>
<td>164.8</td>
<td>–90.1</td>
<td>–70.6</td>
<td>–149.7</td>
<td>44.2</td>
</tr>
<tr>
<td>d(CpG)2</td>
<td>d(CpG)</td>
<td>158.6</td>
<td>–86.4</td>
<td>–65.6</td>
<td>–159.9</td>
<td>48.3</td>
</tr>
<tr>
<td>d(CpG)2</td>
<td>d(CpG)</td>
<td>158.9</td>
<td>–86.0</td>
<td>–67.2</td>
<td>–156.8</td>
<td>47.8</td>
</tr>
<tr>
<td>d(CpG)2</td>
<td>d(CpG)</td>
<td>164.4</td>
<td>–86.7</td>
<td>–76.7</td>
<td>–136.3</td>
<td>41.2</td>
</tr>
<tr>
<td>d(GpG)2</td>
<td>d(GpG)</td>
<td>164.1</td>
<td>–86.2</td>
<td>–77.1</td>
<td>–135.8</td>
<td>41.1</td>
</tr>
<tr>
<td>d(GpG)2</td>
<td>d(GpG)</td>
<td>152.0</td>
<td>–82.6</td>
<td>–63.7</td>
<td>–159.4</td>
<td>47.5</td>
</tr>
<tr>
<td>d(CpC)2</td>
<td>d(CpC)</td>
<td>170.3</td>
<td>–90.8</td>
<td>–69.3</td>
<td>–160.0</td>
<td>46.8</td>
</tr>
<tr>
<td>Av (SD) [⁴°]</td>
<td>159.9 (7)</td>
<td>–76.9 (38)</td>
<td>–67.4 (6)</td>
<td>–157.7 (14)</td>
<td>47.0 (4)</td>
<td>–103.0 (7)</td>
</tr>
<tr>
<td>Exptl (SD) [⁴°]</td>
<td>184 (11)</td>
<td>–95 (10)</td>
<td>–62 (15)</td>
<td>–176 (9)</td>
<td>48 (11)</td>
<td>–102 (14)</td>
</tr>
</tbody>
</table>

[a] Computed at BLYP-D/TZ2P and COSMO for simulating aqueous solution. See Figure 2 for definitions and Figure 3–5 for structures. [b] Single strands that constitute the double strand. [c] Average of the two χ angles present in each dinucleoside monophosphate (see Figure 2). [d] Average with standard deviation (SD) over the ten dinucleoside monophosphate double strands. [e] Experimental values with standard deviation (SD) from X-ray crystal structures of oligodeoxynucleotides in B-DNA conformation (ref. [21]).

### Figure 3. Structures and relative complexation energies ΔE_double (Equation 1 b) of dinucleoside phosphate duplexes Na₂–d(ApA)₂, Na₂–d(ApT)₂ and Na₂–d(TpA)₂, optimized at COSMO/H₂O–BLYP-D/TZ2P (left: major-groove side; right: top view).

<table>
<thead>
<tr>
<th>ds</th>
<th>E [kcal mol⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>d(ApA)₂</td>
<td>7.9 kcal mol⁻¹</td>
</tr>
<tr>
<td>d(ApT)₂</td>
<td>8.5 kcal mol⁻¹</td>
</tr>
<tr>
<td>d(TpA)₂</td>
<td>13.4 kcal mol⁻¹</td>
</tr>
</tbody>
</table>
Table 3. Hydrogen-bond structure (in Å) and stability (in kcal mol$^{-1}$) of the 10 dinucleoside monophosphate B-DNA double strands.\[a\]

<table>
<thead>
<tr>
<th>ds</th>
<th>ss 1/ss 2</th>
<th>(\Delta E_{\text{str}})</th>
<th>(\Delta E_{\text{int}})</th>
<th>(\Delta E_{\text{double}})</th>
<th>(\Delta E_{\text{double,rel}})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\text{N6–O4})</td>
<td>(\text{N1–N3})</td>
<td>(\text{O6–N4})</td>
<td>(\text{N1–N3})</td>
<td>(\text{N2–O2})</td>
</tr>
<tr>
<td>(\text{d(ApG)}_2)</td>
<td>(\text{d(ApA)}_2)</td>
<td>2.88</td>
<td>2.81</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\text{d(TpT)}_2)</td>
<td>(\text{d(ApT)}_2)</td>
<td>2.99</td>
<td>2.77</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\text{d(TpA)}_2)</td>
<td>(\text{d(ApT)}_2)</td>
<td>2.93</td>
<td>2.78</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\text{d(GpA)}_2)</td>
<td>(\text{d(GpA)}_2)</td>
<td>2.93</td>
<td>2.78</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\text{d(ApG)}_2)</td>
<td>(\text{d(ApG)}_2)</td>
<td>2.88</td>
<td>2.80</td>
<td>2.87</td>
<td>2.88</td>
</tr>
<tr>
<td>(\text{d(GpA)}_2)</td>
<td>(\text{d(GpA)}_2)</td>
<td>2.98</td>
<td>2.78</td>
<td>2.81</td>
<td>2.88</td>
</tr>
<tr>
<td>(\text{d(ApC)}_2)</td>
<td>(\text{d(GpC)}_2)</td>
<td>2.88</td>
<td>2.82</td>
<td>2.85</td>
<td>2.87</td>
</tr>
<tr>
<td>(\text{d(GpC)}_2)</td>
<td>(\text{d(GpC)}_2)</td>
<td>2.92</td>
<td>2.78</td>
<td>2.84</td>
<td>2.86</td>
</tr>
<tr>
<td>(\text{d(GpG)}_2)</td>
<td>(\text{d(GpG)}_2)</td>
<td>–</td>
<td>–</td>
<td>2.84</td>
<td>2.90</td>
</tr>
<tr>
<td>(\text{d(CpA)}_2)</td>
<td>(\text{d(GpC)}_2)</td>
<td>–</td>
<td>–</td>
<td>2.84</td>
<td>2.88</td>
</tr>
<tr>
<td>(\text{d(CpG)}_2)</td>
<td>(\text{d(GpG)}_2)</td>
<td>–</td>
<td>–</td>
<td>2.85</td>
<td>2.85</td>
</tr>
<tr>
<td>(\text{AT}_2)</td>
<td>(\text{AT}_2)</td>
<td>2.89</td>
<td>2.80</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\text{GC}_2)</td>
<td>(\text{GC}_2)</td>
<td>2.89</td>
<td>2.81</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\text{AT}^{[c]})</td>
<td>(\text{AT}^{[c]})</td>
<td>2.91</td>
<td>2.82</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\text{GC}^{[d]})</td>
<td>(\text{GC}^{[d]})</td>
<td>–</td>
<td>–</td>
<td>2.85</td>
<td>2.90</td>
</tr>
</tbody>
</table>

[a] Computed at BLYP-D/TZ2P and COSMO for simulating aqueous solution. Parameters of the two base pair dimers (AT)$_2$ and (GC)$_2$ have been calculated for comparison. See Figure 2 for definitions and Figure 3–6 for structures. [b] Single strands that constitute the double strand. [c] See Equations 1 and 3. [d] From ref. [4g].

Thus, we have computed the complexation energy $\Delta E_{\text{double}}$ associated with forming simple (AT)$_2$ stacks and (GC)$_2$ stacks from the corresponding “single strands” of AT + AT stacks and GC + GC stacks, respectively, each of them without the sugar–phosphate–sugar backbone. The equilibrium structures of (AT)$_2$ and (GC)$_2$ are shown in Figure 6, whereas the corresponding hydrogen-bond distances and complexation energies $\Delta E_{\text{double}}$ are reported in Table 3. These backboneless systems were obtained by removing the sugar phosphate units from (d(ApT)$_2$) and (d(GpC)$_2$) substituting C1’ with a H atom, and then optimizing the structure of the stacked base pairs (AT)$_2$ and (GC)$_2$. Likewise, the constituting single strand analogues were obtained by removing, in addition, one of the two AT stacks from (AT)$_2$ and one of the two GC stacks from (GC)$_2$ followed by optimization of the remaining stack of DNA bases. The computed complexation energies $\Delta E_{\text{double}}$ appear to be –19.1 and –25.1 for (AT)$_2$ and (GC)$_2$, respectively. This corresponds to effective Watson–Crick bond strengths per base pair of $\Delta E_{\text{double}}/2 = –9.6$ and –12.6 kcal mol$^{-1}$, respectively. This is almost, that is, within a kcal mol$^{-1}$, the same effective complexation energy as for simple base pairs.

We conclude that the geometrical constraints imposed by the backbone reduce the stability of B-DNA by 0–3 kcal mol$^{-1}$ and 0–2 kcal mol$^{-1}$ per AT and GC base pair, respectively. This weakening shows up in the aforementioned strain energy $\Delta E_{\text{strain}}$, a somewhat weakened interaction energy $\Delta E_{\text{int}}$ but also in slightly (i.e., by a few hundredth of an Å) stretched or compressed hydrogen-bond distances (see Table 3). As soon as the backbone is removed from the double-helical model systems, the effective Watson–Crick complexation energy is enhanced and approaches that of the isolated base pairs. Likewise, hydrogen-bond distances in (AT)$_2$ and (GC)$_2$ relax and essentially adopt the values in the corresponding isolated base pairs (see Table 3).\[4f\] Thus, the values for N6–O4 and N1–N3 distances with an AT base pair of (AT)$_2$ are, on the average, 2.80 and 2.90 Å, respectively, whereas the values for O6–N4, N1–N3 and N2–O2 distances within a GC base pair in (GC)$_2$ amount to 2.84, 2.88 and 2.84 Å, respectively.
major contribution to the overall trend in stability of the structure to the geometry they adopt in the double helix. The required to deform the single strands from their own equilibrium single strands, and the destabilizing strain energy that is restabilizing interaction energy between two complementary single strands. The decomposition of the complexation energy into two counterbalancing contributions: the stranding energy of the model double helix increases as the number of GC pairs increases from zero (−14 to −19 kcal mol\(^{-1}\)) to one (−19 to −23 kcal mol\(^{-1}\)) to two (−24 to −27 kcal mol\(^{-1}\)).

The reason that the hydrogen-bonding interaction \(\Delta E_{\text{int}}\) is weaker in the model double helices is the geometrical constraint imposed by the backbone. In addition to preventing the hydrogen bonds to adopt their optimal Watson–Crick geometry, the deformation of the backbone, upon bending the single strands into the right shape for entering into Watson–Crick complexation, also contributes a strain energy term \(\Delta E_{\text{strain}}\) that further weakens the overall complexation energy \(\Delta E_{\text{double}} = \Delta E_{\text{strain}} + \Delta E_{\text{int}}\) for forming the model double helices. Interestingly, the stability of model double helices with a particular number of AT and GC pairs varies significantly depending on other structural features, namely: (1) the partitioning of purine and pyrimidine bases over the two single strands; and (2) the order in which two DNA bases appear in the 5′→3′ direction along a strand. This trend, which is superimposed on that of the interaction term, stems from a variation in strain energy.

### Conclusions

Our systematic computational analysis of the structure and the stability of all possible single strands and double strands of dinucleoside monophosphates with sodium counter ions in aqueous solution is the first quantum chemical study that copes with these large model systems using dispersion-corrected DFT and implicit solvation (COSMO). Our computations agree well with the available experimental data, notably X-ray crystallographic structures of oligodeoxynucleotides in B-DNA conformation.

We propose a physical model for understanding trends in stability for the formation of double-stranded from single-stranded DNA dimers through a decomposition of the complexation energy into two counterbalancing contributions: the stabilizing interaction energy between two complementary single strands, and the destabilizing strain energy that is required to deform the single strands from their own equilibrium structure to the geometry they adopt in the double helix. The major contribution to the overall trend in stability of the double strand stems from Watson–Crick hydrogen bonding contained in the interaction term \(\Delta E_{\text{int}}\) between the complementary single strands. The effective strength of the hydrogen-bonding interaction \(\Delta E_{\text{int}}\) is somewhat less in the model double helices than in individual base pairs but it still follows the number of hydrogen bonds. Consequently, the stability \(\Delta E_{\text{double}}\) of the model double helix increases as the number of GC pairs increases from zero (−14 to −19 kcal mol\(^{-1}\)) to one (−19 to −23 kcal mol\(^{-1}\)) to two (−24 to −27 kcal mol\(^{-1}\)).

### Acknowledgements

We thank the following organizations for financial support: HPC-EUROPA2 (project number: 211437), with the support of the European Community - Research Infrastructure Action of the FP7 "Co-ordination and support action" program, the National Research School Combination - Catalysis (NRSC-C), and the Netherlands Organization for Scientific Research (NWO-CW and NWO-EW).

**Keywords:** density functional calculations · DNA structures · hydrogen bonds · stacking interactions · Watson–Crick base pairs

---


