

# Stacking with the unnatural DNA base 6-ethynylpyridone

Douglas J. Gibson and Tanja van Mourik\*

EaStCHEM School of Chemistry, University of St Andrews, North Haugh, St Andrews KY16 9ST (UK)

Abstract

It was previously reported that the incorporation of 6-ethynylpyridone (E) into a DNA duplex (replacing T in a T:A base pair) leads to DNA duplexes that are more stable than the T:A-containing duplexes. DFT calculations at the M06-2X/6-31+G(d) and BLYP-D3/6-31+G(d) levels on various base pairs, stacked bases and stacked base pairs in continuum solvation water suggest that the observed increased stability of E:A-containing duplexes is due to the combined effects of stronger base pairing and enhanced stacking of the E:A base pair.

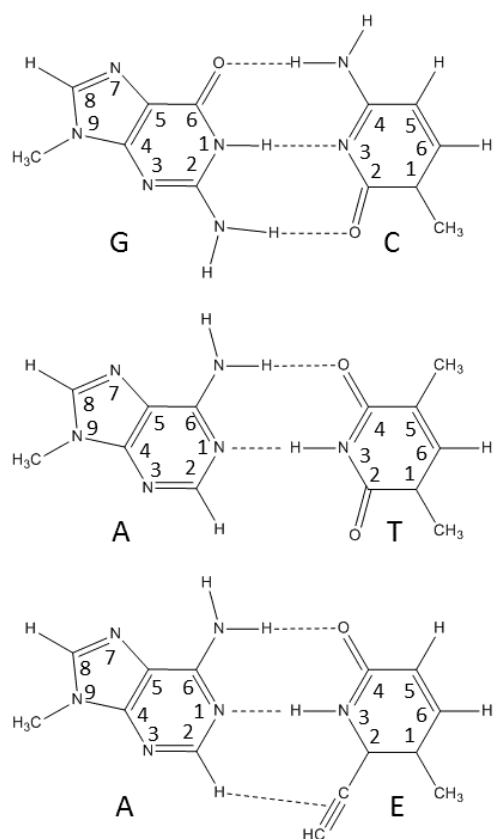
**Keywords:** Unnatural base pair, 6-ethynylpyridone, density functional theory, stacking, M06-2X, BLYP-D3

## 1. Introduction

DNA is one of life's most important molecules, as it carries all genetic information. DNA uses a four-letter alphabet to encode genetic instructions: A (adenine), T (thymine), G (guanine) and C (cytosine). T selectively pairs with A and C with G, *via* specific hydrogen bonds. Cytosine-guanine (C:G) base pairs have three hydrogen bonds (H-bonds), whereas thymine-adenine (T:A) base pairs have only two (see Scheme 1); C:G base pairs are therefore more stable than T:A base pairs. Specific pairing is essential for the correct transmission of the genetic information through DNA replication (the process of producing two identical copies of DNA from one original DNA molecule), transcription (where a DNA segment is copied into messenger RNA – mRNA) and translation (decoding of mRNA produced by transcription to produce a chain of amino acids). Over the last few decades, nucleotide analogues have been used to increase the DNA alphabet, forming unnatural base pairs. This increases the genetic information that can be encoded by DNA. A plethora of studies have explored extending the genetic lexicon by incorporating unnatural base pairs into DNA or RNA [1-10]. Most unnatural base pairs successfully incorporated into DNA or RNA contain different hydrogen-bond patterns than the canonical base pairs, though also base pairs with no hydrogen bonds at all have been employed [11-14].

---

\* Corresponding author. Email address: [tanja.vanmourik@st-andrews.ac.uk](mailto:tanja.vanmourik@st-andrews.ac.uk); Phone: +44 (0)1334 463822



**Scheme 1.** Atom numbering in the G:C, A:T and A:E base pairs, consisting of the methylated forms of the bases.

6-Ethynylpyridone (E), a thymine analogue, was recently reported to bind to adenine with greater selectivity and affinity than thymine itself, even when incorporated into a DNA duplex [15]. Compared to T:A, an E:A base pair contains one additional intermolecular interaction, a C-H $\cdots$  $\pi$  interaction (Scheme 1). The increased stability when replacing a T:A pair in the examined DNA duplex (5'-TCTTTCXTTCT-3':3'-GAAAAGYAAGAA-5') by an E:A pair is of similar magnitude as replacing the T:A pair with a C:G pair [15]. This was deduced by measurements of UV melting temperature, which showed that the replacement of a T:A pair with an E:A pair increased the melting point of the DNA duplex by 2.6 °C, comparable to the observed increase (2.8 °C) when the T:A pair was replaced by a C:G pair. These observations indicate that the stability of the E:A base pair almost equals that of a C:G pair. A subsequent DFT study found that E does indeed bind with adenine with more affinity than thymine does; however, the calculations showed that the stability of the E:A pair is intermediate between those of the canonical T:A and C:G base pairs [16]. Thus, the computed stability of E:A does not fully explain the remarkable increased stability (similar to the C:G base pair) observed in the experiments. This implies that the increased stability of E:A observed experimentally cannot be explained fully by base-pair interactions, and that other interactions must also contribute. One hypothesis put forward is that the increased stability observed is due to enhanced stacking interactions within the DNA duplex structure [16]. We also note here that the stability of DNA is largely due to stacking interactions, rather than H-bonding interactions [17].

It seems therefore reasonable to assume that the increased stability of duplexes with T:A replaced by E:A is due to greater stacking.

In the current manuscript we investigate using dispersion-corrected density functional theory whether an E:A base pair stacks more strongly with A:T or C:G than other base pairs. We find that increased stacking is not evident when considering stacking of individual bases. However, calculations on stacked base pairs provide indication of increased stacking of E:A compared to the canonical base pairs.

## 2. Methodology

*Stacking orientation:* Due to their low symmetry, DNA bases have two distinguishable faces [18]. Rose *et al.* [19] proposed a naming convention for the different faces of ring compounds: if the direction of ascending atom numbers in the ring is clockwise, the face is labelled  $\alpha$ ; if it is counterclockwise, it is  $\beta$ . Scheme 1 shows the atom labelling in the DNA bases as well as 6-ethynylpyridone. Note that in the following, A, T, C, G and E refers to the methylated forms of the bases (9-methylation for the purines and 1-methylation for the pyrimidines and E). The atom numbering of 6-ethynylpyridone is decided in correspondence to thymine (taken into account how both bases bind to adenine). In B-DNA, Watson-Crick pairing between a pyrimidine and purine base occurs as  $\alpha$  with  $\alpha$  or  $\beta$  with  $\beta$ . When looking down the 5'-strand in the antiparallel arrangement of the two strands in double-helical DNA, the purines show their  $\beta$ -faces and the pyrimidines their  $\alpha$ -faces. Adjacent pyrimidine bases therefore stack with the  $\beta$ -face of the 5'-base on the  $\alpha$ -face of the 3'-base ( $\beta/\alpha$  stacking). Adjacent purine bases stack  $\alpha/\beta$ , whereas a 5'-pyrimidine/3'-purine pair stacks  $\beta/\beta$  and a 5'-pyrimidine/3'-purine pair stacks  $\alpha/\alpha$ . Previously, we had denoted these as face-to-face ( $\alpha/\alpha$  or  $\beta/\beta$ ) or face-to-back ( $\alpha/\beta$  or  $\beta/\alpha$ ) in stacked DNA base pairs [20-22]. Note that if there is no information on which base is on the 3' or 5' side (like in isolated stacked base pairs), then there is no distinction between  $\alpha/\alpha$  and  $\beta/\beta$  or between  $\alpha/\beta$  and  $\beta/\alpha$ .

*H-bonded base pairs:* The geometries of the methylated bases adenine (A), guanine (G), thymine (T) and cytosine (C), and the unnatural base 6-ethynylpyridone (E) were optimised with the M06-2X density functional [23] and the 6-31+G(d) basis set. The calculations were done in the gas phase and in aqueous solution, modelled by the Polarizable Continuum Model (PCM) [24] using the integral equation formalism variant (IEFPCM). All calculations were performed with the Gaussian 09 package [25] and employed Gaussian's "ultrafine" integration grid. Base pairs were then constructed by placing the bases in favourable base-pair positions using the visualisation program GDIS [26]. The bases were placed such that the bonds that model the link to the sugar moiety in DNA (N3-C<sub>methyl</sub> in the purines; N1-C<sub>methyl</sub> in the pyrimidines; C1-C<sub>methyl</sub> in E) are in a suitable position for linking to the sugar-phosphate backbone in DNA (pointing down-left in the left base and down-right in the right base - see Scheme 1). We considered the natural base pairs (G:C and A:T), the mispairs containing T (T:C, T:G and T:T) and the unnatural base pairs consisting of E paired with the four natural bases (E:A, E:C, E:G and E:T). The base pairs were optimised with M06-2X/6-31+G(d) in the gas phase

and in aqueous solution. The gas-phase interaction energies were corrected for basis set superposition error (BSSE) using the counterpoise procedure [27], invoked by Gaussian's "Counterpoise" keyword. As it is not possible to implement counterpoise during a PCM calculation with Gaussian, the interaction energies in water were corrected using the BSSE calculated from the gas-phase calculation:

$$\Delta E(\text{PCM}) = E_{\text{AB}}^{\{\text{AB}\}}(\text{AB})(\text{PCM}) - \text{BSSE}(\text{gas}) - E_{\text{A}}^{\{\text{A}\}}(\text{A})(\text{PCM}) - E_{\text{B}}^{\{\text{B}\}}(\text{B})(\text{PCM}) \quad (1)$$

$$\text{BSSE}(\text{gas}) = E_{\text{A}}^{\{\text{AB}\}}(\text{AB})(\text{gas}) + E_{\text{B}}^{\{\text{AB}\}}(\text{AB})(\text{gas}) - E_{\text{A}}^{\{\text{A}\}}(\text{A})(\text{gas}) - E_{\text{B}}^{\{\text{B}\}}(\text{B})(\text{gas}) \quad (2)$$

In eqs. (1) and (2) A and B refer to the two separate fragments (individual bases, in this case); the superscripts in curly brackets indicate the basis set (monomer basis set {A} or {B}, or the dimer basis set {AB}); subscripts denote the molecular system; and systems in round brackets denote the geometry employed for the energy calculation (dimer geometry (AB) or optimised monomer geometry (A) or (B)).

*Stacked bases:* We considered T stacked with the four natural bases as well as E, and E stacked with the five bases considered. As isolated stacks preferably adopt relative orientations (twist angles) that are not feasible in DNA [20-22], we first optimised the structures of the corresponding dinucleotides (the stacked bases connected by the sugar-phosphate backbone). We considered dinucleotides with T or E at the 3'-end and one of the five bases at the 5'-end as well as dinucleotides with T or E at the 5'-end and the other base at the 3'-end. After optimisation of the dinucleotides, the sugar-phosphate backbone was removed and a methyl group added at the N1-position in C or T, the N9-position in A and G or the C5 position in E. The positions of the atoms belonging to the methyl group were optimised using M06-2X/6-31+G(d), keeping the rest of the base at the dinucleotide geometry. The stacking energy was then calculated at the same level of theory as for the H-bonded base pairs (CP-corrected M06-2X/6-31+G(d)/PCM).

*Stacked base pairs:* Like for the stacked bases, we initially included the sugar-phosphate backbone. For the stacked base pairs, we adopted the numbering scheme of a base pair step following Svozil *et al.* [28]. The stacked base pairs are denoted 12/34, where 12 and 34 are base pairs. Base 1 is stacked above 3 and 2 above 4, with bases 1 and 4 at the 5' end and 3 and 2 at the 3'-end. We considered CG/XY and XY/AT, with XY replaced by TA, EA and CG. The starting structures were constructed from the corresponding optimised dinucleotides, placed in favourable base-pair positions in duplex conformation using GDIS. The dinucleotide duplexes were optimised with BLYP-D3 (BLYP augmented with Grimme's D3 dispersion correction [29]) in conjunction with the TZVP basis set [30] and PCM water. As for the stacked bases, the sugar-phosphate backbones were then removed, methyl groups were added and the positions of the methyl atoms optimised with M06-2X/6-31+G(d). The stacking interaction was then calculated at the same level of theory as for the H-bonded base pairs and stacked bases (CP-corrected M06-2X/6-31+G(d)/PCM), taking the H-bonded base pairs as the fragments.

*SAPT*: The stacked base pairs were also analysed with Symmetry-Adapted Perturbation Theory (SAPT) [31]. SAPT calculates the interaction energy directly without computing the dimer and monomer energies. We employed a scaled version, sSAPT0 [32], of the simplest truncation of the SAPT expansion (SAPT0), with a truncated aug-cc-pVDZ basis set labelled jun-cc-pVDZ (previously called aug-cc-pVDZ' [33]), using the program Psi4 [34]. The sSAPT0/jun-cc-pVDZ level of theory was recommended as an efficient, good-performing level of SAPT [32]. For comparison with the SAPT results, “vertical” gas-phase interaction energies were computed as well:

$$\Delta E^{\text{vert}}(\text{gas}) = E_{\text{AB}}^{\{\text{AB}\}}(\text{AB})(\text{gas}) - E_{\text{A}}^{\{\text{AB}\}}(\text{A})(\text{gas}) - E_{\text{B}}^{\{\text{AB}\}}(\text{B})(\text{gas}) \quad (3)$$

These correspond to gas-phase counterpoise-corrected interaction energies without inclusion of monomer deformation energies.

*Geometrical analysis*: We calculated the DNA base-step parameters Tilt, Rise and Twist using an in-house Fortran program, described in Ref. [21]. The Tilt is the angle between the two (H-bonded or stacked) base planes, the Twist (for stacked bases) is the rotation of the upper base with respect to the lower, whereas the Rise is the vertical distance between two stacked bases. A detailed description how these parameters are calculated can be found in the Supplementary material.

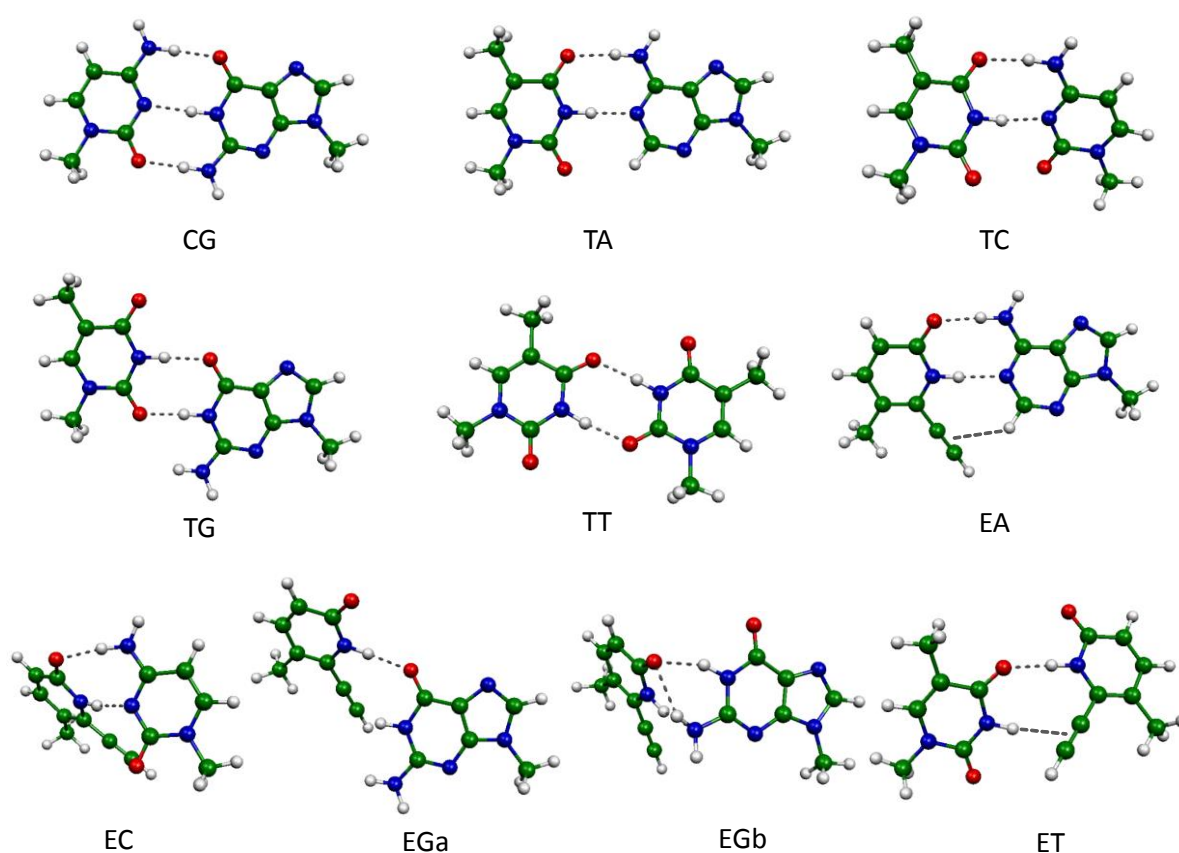
*Optimised structures*: The Cartesian coordinates of all optimised base pairs, dinucleotides, and dinucleotide duplexes, as well as the stacked base pairs extracted from the optimised dinucleotides and the stacked base pairs extracted from the optimised dinucleotide duplexes can be found in the Supplementary material.

### 3. Results and discussion

#### 3.1 H-bonded base pairs

The structures of the optimised H-bonded base pairs are shown in Figure 1. Table 1 compares the CP-corrected M06-2X/6-31+G(d)/PCM interaction energies of the H-bonded base pairs with the CP-corrected MP2/aug-cc-pVDZ single-point energies by Halder *et al.* [16]. The M06-2X and MP2 results are in close agreement. The largest deviation occurs for E:A, which MP2 predicts to be 5 kJ/mol more stable than M06-2X. The tilt values listed in Table 1 show that the natural base pairs (A:T and G:C), as well as the T:G and T:T mispairs, are very nearly planar. Other pyrimidine:pyrimidine mispairs and all base pairs involving E are considerably non-planar. Note that we found two distinct E:G base pairs: E:Ga, which corresponds to the base pair optimised by Halder *et al.*, and E:Gb. E:Gb is the more stable, but also the most non-planar, of the two. Overall, the C:G base pair is the most stable base pair as it is stabilised by three H-bonds. The E:A base pair is more stable than the natural T:A base pair. This is partially due to the C-H••• $\pi$  interaction present in E:A but not in T:A. Halder *et al.* used AIM (Atoms in Molecules) analysis to show that the C-H••• $\pi$  interaction in E:A and

the N-H... $\pi$  interaction in E:Ga (but not the C-H...O interaction in T:A) can be classified as H-bonds [16]. In addition, NBO (Natural Bond Orbital) analysis and IR frequency data suggest the N-H...O interaction is stronger in E:A than in T:A [16]. However, the E:A base pair is considerably less stable than C:G (which is stabilised by three conventional H-bonds). Therefore, the increased stability of E:A over T:A alone cannot explain the increased thermostability observed experimentally when a T:A pair is replaced by an E:A pair in DNA duplexes [35]. It was suggested that base pairing is not the only source of the increased stability observed in DNA duplexes containing EA pairs, but that enhanced stacking, facilitated by the conjugation of acetylenic  $\pi$  electrons with the  $\pi$  system of the ring, may contribute to their greater stability [16]. In the next section we compare the stacking strength of E with those of the other bases.



**Fig. 1.** Base-pair structures optimised at the M06-2X/6-31+G(d)/PCM level of theory.

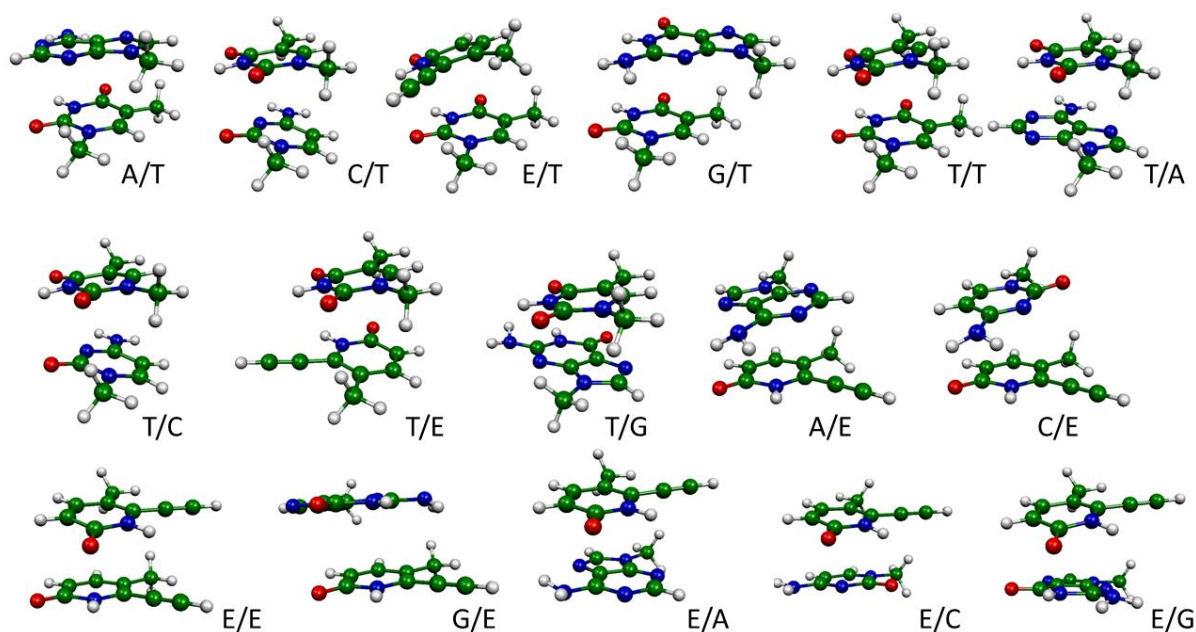
**Table 1.**Interaction energies (in kJ/mol) of the H-bonding base pairs in water<sup>a</sup>

Base pair	$\Delta E(\text{PCM})$		Tilt <sup>c</sup>
	M06-2X/6-31+G(d)	MP2/aug-cc-pVDZ <sup>b</sup>	
C:G	-62.4	-61.1	2
T:A	-37.6	-39.7	0
T:C	-34.6	-33.9	47
T:G	-40.5	-41.8	3
T:T	-36.0	-38.1	0
E:A	-40.9	-46.0	21
E:C	-38.0	-40.6	60
E:Ga	-23.3	-21.3	41
E:Gb	-34.7	n/a	71
E:T	-22.3	-23.0	41

<sup>a</sup> All interaction energies are CP-corrected.<sup>b</sup> Taken from Ref. [16]. Calculated at M05-2X/6-31+G(d,p) optimised geometries. Solvent modelled using CPCM.<sup>c</sup> Using the M06-2X/6-31+G(d)/PCM optimised geometries.

### 3.2 Stacked bases

The interaction energies, tilt and rise of the stacked bases are listed in Table 2. The corresponding structures are given in Figure 2. The mixed purine/pyrimidine stacks have tilt values close to 180° because they stack  $\alpha/\alpha$  or  $\beta/\beta$ . The purine/purine and pyrimidine/pyrimidine pairs have face-to-back stacking ( $\alpha/\beta$  or  $\beta/\alpha$ ) and therefore have tilt values close to 0°. With a tilt value of 155°, E/G exhibits the largest deviation from the ideal structure of two parallel bases. It also has the largest rise (3.58 Å). All other stacks have a rise value around 3.2 – 3.3 Å. This is slightly shorter than the average distance between consecutive bases in B-DNA (3.4 Å [36]), but about 0.1-0.2 Å longer than in stacks optimised at the same level of theory but in the absence of the sugar-phosphate backbone [20-22]. In some stacks (for example A/T), replacing T by E leads to a slight increase in stacking energy. However, in other cases (for example C/T) such a replacement causes a decrease in stacking energy. Overall there is no indication that E stacks better than T in isolated stacks.



**Fig. 2.** Stacked bases optimised at the M06-2X/6-31+G(d)/PCM level of theory. The 5'-base is always on top.

**Table 2.**

Stacking interaction energies (in kJ/mol) of the stacked bases in water, calculated at the CP-corrected M06-2X/6-31+G(d)/PCM level. Rise (in Å) and Tilt (in degrees) are given as well.

5'-X/Y-3'	$\Delta E(\text{PCM})$	Rise	Tilt	5'-X/Y-3'	$\Delta E(\text{PCM})$	Rise	Tilt
A/T	-30.9	3.19	174	A/E	-32.4	3.20	173
C/T	-21.6	3.30	12	C/E	-19.7	3.34	14
E/T	-25.2	3.24	19	E/E	-26.7	3.33	16
G/T	-20.8	3.26	173	G/E	-19.8	3.25	174
T/T	-24.4	3.21	15	T/E	-23.4	3.31	16
T/A	-30.5	3.19	173	E/A	-27.7	3.30	171
T/C	-23.0	3.24	16	E/C	-22.1	3.27	17
T/E	-23.4	3.31	16	E/T	-25.2	3.24	19
T/G	-28.4	3.19	170	E/G	-27.3	3.58	155

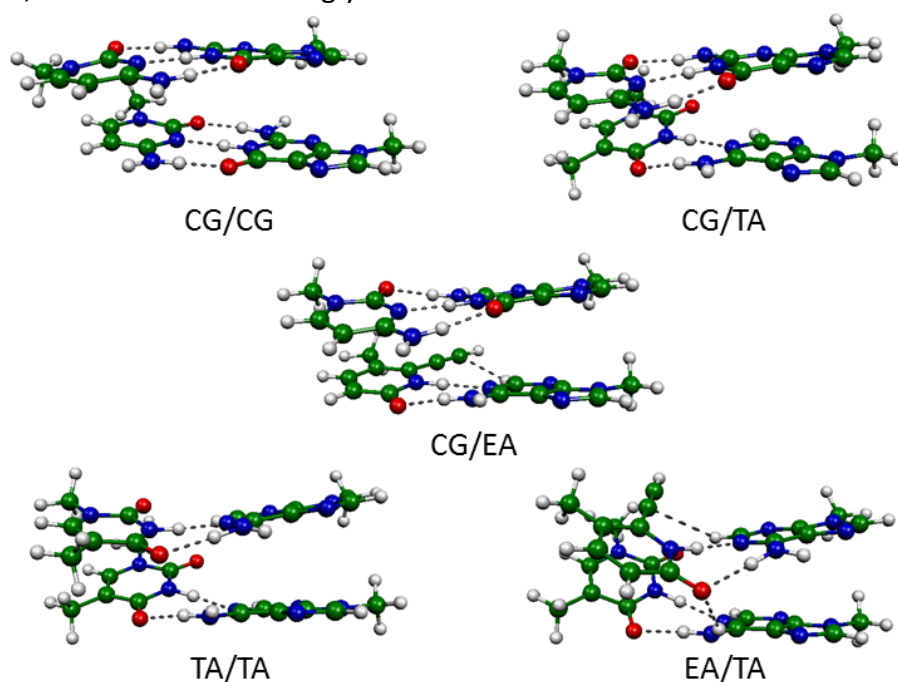
### 3.3 Stacked base pairs

Table 3 shows the stacking interaction energies (in kJ/mol) of the stacked base pairs. The structures are displayed in Figure 3. The first three rows show systems with a 5'-C:G-3' base pair stacking with 3'-X:Y-5' (X:Y = C:G, T:A and E:A, respectively), whereas the last three rows show systems with a 3'-T:A-5' base pair stacking with 5'-X:Y-3' (X:Y = C:G, T:A and E:A, respectively). This conforms to the duplex studied by Halder *et al.*, where X:Y is located



between a 5'-C:G-3' and 3'-T:A-5' base pair. Note that the 5'C/3'T:5'A/G3' system is shown twice in Table 3 to aid analysis. Note that the carbonyl oxygen in EA/TA forms two hydrogen bonds; one of these links the E unnatural base in the upper base pair with the A base in the lower base pair. This H-bond may be responsible for the large stacking interaction in this stacked base pair.

Considering the base pairs stacking with 5'-C:G-3', replacing the 3'-C:G-5' base with 3'-E:A-5' yields significantly increased stabilisation (from -2.46 to -20.00 kJ/mol, using the PCM numbers). However, replacing 3'-T:A-5' by 3'-E:A-5' yields a slight destabilisation (from -23.30 to -20.00 kJ/mol). Inspection of the last three rows in Table 3 shows that, with the 3'-T:A-5' base pair, E:A stacks more strongly than either C:G or A:T.



**Fig. 3.** Stacked base pairs optimised at the BLYP-D3/6-31+G(d)/PCM level of theory. The 5'-base of the left and right base pair are at the top left and bottom right, respectively.

**Table 3.**

Stacking interaction energies (in kJ/mol) of the stacked base pairs in water and gas phase, calculated at different levels of theory.<sup>a</sup>

Stacked base pair	$\Delta E(\text{gas})$	$\Delta E(\text{PCM})$	$\Delta E^{\text{vert}}(\text{gas})$	sSAPTO
CG/CG	7.69	-2.46	-43.85	-49.86
CG/TA	-22.26	-23.30	-63.64	-71.91
CG/EA	-16.45	-20.00	-52.26	-65.10
CG/TA	-22.26	-23.30	-63.64	-71.91
TA/TA	-25.32	-24.80	-65.73	-76.88
EA/TA	-23.89	-27.94	-69.41	-84.57

<sup>a</sup> The interaction energies are calculated taking the H-bonded base pairs as the fragments.

**Table 4.**

Base-step parameters (Rise in Å; Tilt in degrees) of the stacked base pairs in water calculated at the CP-corrected M06-2X/6-31+G(d)/PCM level.<sup>a</sup>

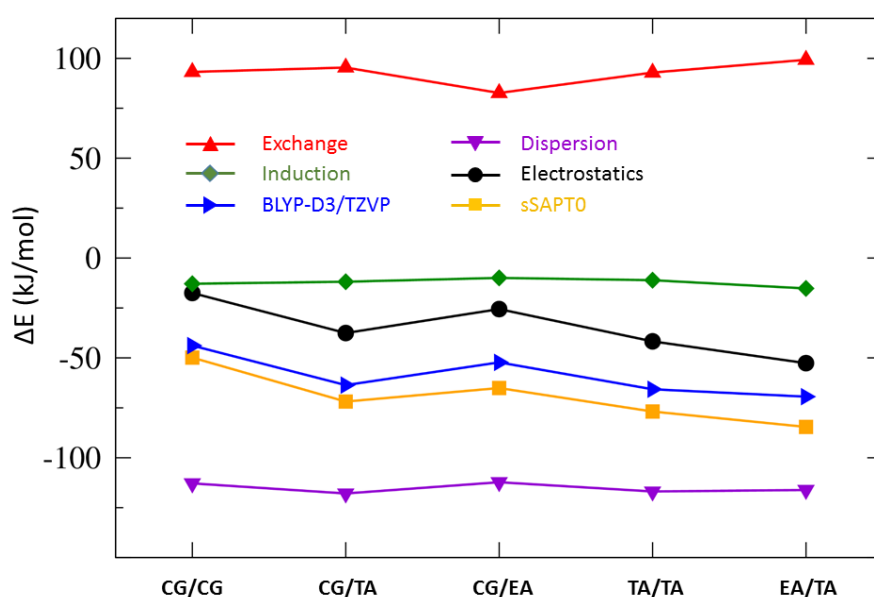
Stacked base pair	Tilt (up)	Tilt (low)	Rise (l)	Tilt (l)	Twist (l)	Rise (r)	Tilt (r)	Twist (r)
CG/CG	12	9	3.24	8	37	3.34	2	46
CG/TA	10	19	3.30	10	38	3.45	4	34
CG/EA	18	13	3.42	7	26	3.47	5	32
TA/TA	24	30	3.30	12	37	3.50	10	37
EA/TA	35	41	3.70	14	47	3.47	10	41

<sup>a</sup> For a stacked base pair 12/34, “up” (upper) and “low” (lower) refer to the H-bonded base pair 1:2 and 3:4, whereas “l” (left) and “r” (right) refer to the stacked bases 1/3 and 2/4.

Table 4 lists the base step parameters for the separate fragments (H-bonded base pairs and stacked bases) of the stacked base pair structures. In an idealised DNA helix, the Twist is 36°, the average distance between two adjacent base pairs (Rise) is 3.4 Å and the Tilt is 0° (conforming to planar and parallel base pairs). However, DNA is very dynamic and it has been shown by analysis of a database consisting of 40 DNA oligomers that significant variations from the idealised values can occur [36]. The Twist angle can vary from as little as 10° to as much as 60°, whereas the Tilt can vary by -25 to +25°. Most of the Rise values calculated for the stacked base pairs are close to the idealised value. The largest deviation occurs for EA/TA (Rise of 3.7 Å for the E/T stack). The Twist angles in our optimised structures vary from 26 to 46° and are all in the accepted range. The Tilt in the stacked bases (left and right) are all within the range observed in the database oligomers, whereas the Tilt in the H-bonded base pairs (upper and lower) exceeds this range in some cases. Again, the largest deviation from the idealised value occurs for the H-bonded base pairs in EA/TA (Tilt values of 35 and 41° for E:A and T:A, respectively). This appears to be due to the preference of E to form non-planar base pairs (see also Table 1); the Tilt of the T:A base pair stacking with E:A is also large (despite

isolated A:T having a Tilt of 0° - see Table 1) as it attempts to keep good stacking with the E:A base pair (which can also be deduced from the relatively small Tilt values for the left and right stacked bases). It thus appears that stacking interactions have a larger effect on nucleotide geometry than H-bonding interactions. The Tilt values in EA/TA are however not much different from those in TA/TA. We also note that the values for CG/EA are in line with those of the other base pairs. We therefore expect that the distortion caused by incorporation of E:A into DNA will likely be minimal.

Also shown in Table 3 are SAPT interaction energies. As the SAPT results do not take into account deformation energies, these are best compared to the vertical CP-corrected gas-phase interaction energies,  $\Delta E^{\text{vert}}(\text{gas})$ , also listed in Table 3. The SAPT results are somewhat larger than the DFT values, but generally follow the same trend. SAPT also provides a decomposition of the interaction energy in physically meaningful components (electrostatic, exchange, induction and dispersion interactions). Figure 4 shows the individual components for the five stacked base pairs considered in this work. Dispersion is clearly the largest component. The electrostatic term is mainly responsible for the order of stability of the different structures; the other terms show smaller variability for the different stacked base pairs. Thus, the electrostatic term is largely responsible for the increased stacking of E:A with T:A. The increased electrostatic term may be due to the hydrogen bond formed between the carbonyl oxygen on E in the upper base pair and an amino hydrogen on A in the lower base pair (see also Figure 3).



**Fig. 4.** Energy components for the stacked base pairs considered in this work calculated at the SAPT0/jun-cc-pVDZ level.

#### 4. Conclusions

To investigate the increased stability previously reported for a DNA duplex, following replacement of a T:A base pair by an E:A base pair (where E is the thymine analogue 6-

ethynylpyridone), we studied base-pairing and stacking interaction energies in various H-bonded base pairs, stacked bases and stacked base pairs, at the counterpoise-corrected M06-2X/6-31+G(d) and BLYP-D3/6-31+G(d) levels of theory in PCM continuum solvation water. Whereas E does bind stronger to A than T does, the increased stability of the H-bonded E:A base pair compared to T:A is insufficient to explain the observed increase in melting temperature of the previously investigated DNA duplex, confirming previously reported results. To investigate whether the increased stability is caused by enhanced stacking of the E:A base pair, we calculated stacking energies of stacked bases and stacked base pairs. Calculations on the stacked bases do not reveal any indication that E stacks stronger than T. However, calculations on the stacked base pairs show that 5'-E:A-3' binds convincingly stronger with a 3'-T:A-5' base pair than the canonical base pairs. There is also some indication that 3'-E:A-5' stacks stronger with a 5'-C:G-3' base pair. SAPT analysis shows that the increased stacking of E:A with T:A is mainly caused by the electrostatic term.

The calculations show that the unnatural E:A base pair experiences increased H-bonding interaction as compared to T:A, as well as increased stacking interaction with adjacent base pairs. Whereas this may aid to explain the increased melting temperature observed experimentally upon replacement of an T:A base pair by an E:A base pair, we would like to point out that the calculations and experiment cannot be directly compared. The DNA melting temperatures are determined by free energies, which result from a concoction of numerous energy contributions, not just stacking and base pairing interactions, and can be context dependent [37,38]. In addition, the calculations and experiments are done with respect to different reference states. DNA melting temperature relates to the temperature at which DNA denatures into single strands, in which probably still some stacking is present, whereas the computations use reference states with no interaction between the bases. However, we do believe the results shed light on aspects of the experimental findings and contribute to a more complete understanding of the effect of substitution of T:A by E:A.

### **Acknowledgements**

We are grateful to EaStCHEM for computational support via the EaStCHEM Research Computing Facility.

### **Appendix A. Supplementary material.**

Supplementary material can be found in the online version, at ....

### **References**

- [1] C. Switzer, S.E. Moroney, S.A. Benner, *J. Am. Chem. Soc.* 111 (1989) 8322-8323.
- [2] J.A. Piccirilli, S.A. Benner, T. Krauch, S.E. Moroney, S.A. Benner, *Nature* 343 (1990) 33-37.
- [3] C.Y. Switzer, S.E. Moroney, S.A. Benner, *Biochemistry* 32 (1993) 10489-10496.
- [4] M.L. Collins, B. Irvine, D. Tyner, E. Fine, C. Zayati, C.-a. Chang, T. Horn, D. Ahle, J. Detmer, L.-P. Shen, J. Kolberg, S. Bushnell, M.S. Urdea, D.D. Ho, *Nucleic Acids Res.* 25 (1997) 2979-2984.

- [5] S.C. Johnson, C.B. Sherrill, D.J. Marshall, M.J. Moser, J.R. Prudent, *Nucleic Acids Res.* 32 (2004) 1937-1941.
- [6] A.M. Sismour, S.A. Benner, *Nucleic Acids Res.* 33 (2005) 5640-5646.
- [7] M. Kimoto, T. Mitsui, Y. Harada, A. Sato, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* 35 (2007) 5360-5369.
- [8] Z. Yang, A.M. Sismour, P. Sheng, N.L. Puskar, S.A. Benner, *Nucleic Acids Res.* 35 (2007) 4238-4249.
- [9] M.M. Georgiadis, I. Singh, W.F. Kellett, S. Hoshika, S.A. Benner, N.G.J. Richards, *J. Am. Chem. Soc.* 137 (2015) 6947-6955.
- [10] L. Zhang, Z. Yang, K. Sefah, K.M. Bradley, S. Hoshika, M.-J. Kim, H.-J. Kim, G. Zhu, E. Jiménez, S. Cansiz, I.T. Teng, C. Champanhac, C. McLendon, C. Liu, W. Zhang, D.L. Gerloff, Z. Huang, W. Tan, S.A. Benner, *J. Am. Chem. Soc.* 137 (2015) 6734-6737.
- [11] D.L. McMinn, A.K. Ogawa, Y. Wu, J. Liu, P.G. Schultz, F.E. Romesberg, *J. Am. Chem. Soc.* 121 (1999) 11585-11586.
- [12] E.L. Tae, Y. Wu, G. Xia, P.G. Schultz, F.E. Romesberg, *J. Am. Chem. Soc.* 123 (2001) 7439-7440.
- [13] D.A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J.M. Foster, I.R. Correa, F.E. Romesberg, *Nature* 509 (2014) 385-388.
- [14] S. Jahiruddin, A. Datta, *J. Phys. Chem. B* 119 (2015) 5839-5845.
- [15] M. Minuth, C. Richert, *Angew. Chem. Int. Ed.* 52 (2013) 10874-10877.
- [16] A. Halder, A. Datta, D. Bhattacharyya, A. Mitra, *J. Phys. Chem. B* 118 (2014) 6586-6596.
- [17] P. Yakovchuk, E. Protozanova, M.D. Frank-Kamenetskii, *Nucleic Acids Res.* 34 (2006) 564-574.
- [18] J. Chen, B. Kohler, *J. Am. Chem. Soc.* 136 (2014) 6362-6372.
- [19] I.A. Rose, K.R. Hanson, K.D. Wilkinson, M.J. Wimmer, *Proc. Natl. Acad. Sci. USA* 77 (1980) 2439-2441.
- [20] L.F. Holroyd, T. van Mourik, *Theor. Chem. Acc.* 133 (2014) 1431-1443.
- [21] R.S. Hunter, T. van Mourik, *J. Comput. Chem.* 33 (2012) 2161-2172.
- [22] T. Mourik, S.W.L. Hogan, *Struct. Chem.* (2015) 145-158.
- [23] Y. Zhao, D.G. Truhlar, *Theor. Chem. Acc.* 120 (2008) 215-241.
- [24] S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* 55 (1981) 117-129.
- [25] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09. Gaussian, Inc., Wallingford, CT, 2009.
- [26] S. Fleming, A. Rohl, *Z. Kristallogr.* 220 (2005) 580-584.
- [27] S.F. Boys, F. Bernardi, *Mol. Phys.* 19 (1970) 553-566.
- [28] D. Svozil, P. Hobza, J. Šponer, *J. Phys. Chem. B* 114 (2010) 1191-1203.
- [29] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* 132 (2010) 154104 (154119 pages).
- [30] A. Schaefer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* 100 (1994) 5829-5835.
- [31] B. Jeziorski, R. Moszynski, K. Szalewicz, *Chem. Rev.* 94 (1994) 1887-1930.
- [32] T.M. Parker, L.A. Burns, R.M. Parrish, A.G. Ryno, C.D. Sherrill, *J. Chem. Phys.* 140 (2014) 094106.
- [33] E.G. Hohenstein, R.M. Parrish, C.D. Sherrill, J.M. Turney, H.F. Schaefer, *The Journal of Chemical Physics* 135 (2011) 174107.

- [34] J.M. Turney, A.C. Simmonett, R.M. Parrish, E.G. Hohenstein, F. Evangelista, J.T. Fermann, B.J. Mintz, L.A. Burns, J.J. Wilke, M.L. Abrams, N.J. Russ, M.L. Leininger, C.L. Janssen, E.T. Seidl, W.D. Allen, H.F. Schaefer, R.A. King, E.F. Valeev, C.D. Sherrill, T.D. Crawford, *WIREs Comput. Mol. Sci.* 2 (2012) 556.
- [35] M. Minuth, C. Richert, *Angew. Chem. Int. Ed.* 52 (2013) 10874-10877.
- [36] M.A. El Hassan, C.R. Calladine, *Phil. Trans. Roy. Soc. London A* 355 (1997) 43-100.
- [37] M. Krepl, M. Otyepka, P. Banáš, J. Šponer, *J. Phys. Chem. B* 117 (2013) 1872-1879.
- [38] J. Šponer, J.E. Šponer, A. Mládek, P. Banáš, P. Jurečka, M. Otyepka, *Methods* 64 (2013) 3-11.