PCCP

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: M. Chawla, A. Poater, P. Besalu-Sala, K. Kalra, R. Oliva and L. Cavallo, *Phys. Chem. Chem. Phys.*, 2018, DOI: 10.1039/C7CP07656H.

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [author guidelines](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the ethical guidelines, outlined in our [author and reviewer resource centre](#), still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Theoretical Characterization of Sulfur-to-Selenium Substitution in an Emissive RNA Alphabet: Impact on H-bonding Potential and Photophysical Properties

Mohit Chawla,^{1,*} Albert Poater,^{1,2} Pau Besalú-Sala,² Kanav Kalra,¹ Romina Oliva,³ and Luigi Cavallo^{1,*}

¹King Abdullah University of Science and Technology (KAUST), Physical Sciences and Engineering Division, KAUST Catalysis Center, Thuwal 23955-6900, Saudi Arabia. ²Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, M^a Aurèlia Campmany, 17003 Girona, Catalonia, Spain. ³Department of Sciences and Technologies, University Parthenope of Naples, Centro Direzionale Isola C4, I-80143, Naples, Italy.

Email: mohitchawla.bt@gmail.com; luigi.cavallo@kaust.edu.sa

Abstract

We employ density functional theory (DFT) and time-dependent DFT (TDDFT) calculations to investigate the structural, energetic and optical properties of a new computationally designed RNA alphabet, where the nucleobases, ^{ts}A, ^{ts}G, ^{ts}C, and ^{ts}U (ts-bases), have been derived by replacing sulfur with selenium in the previously reported tz-bases, based on the isothiazolo[4.3-d]pyrimidine heterocycle core. We find out that the modeled non-natural bases have minimal impact on the geometry and energetics of the classical Watson-Crick base pairs, thus potentially mimicking the natural bases in a RNA duplex in terms of H-bonding. In contrast, our calculations indicate that H-bonded base pairs involving the Hoogsteen edge of purines are destabilized as compared to their natural counterparts. We also focus on the photophysical properties of the non-natural bases and correlate their absorption/emission peaks to the strong impact of the modification on the energy of the lowest unoccupied molecular orbital. It is indeed stabilized by roughly 1.1-1.6 eV as compared to the natural analogues, resulting in a reduction of the gap between the highest occupied and the lowest unoccupied molecular orbital from 5.3-5.5 eV in the natural bases to 3.9-4.2 eV in the modified ones, with a consequent bathochromic shift in the absorption and emission spectra. Overall, our analysis

clearly indicates that the newly modelled ts-bases are expected to exhibit better fluorescent properties as compared to the previously reported tz-bases, while retaining similar H-bonding properties. In addition, we show that a new RNA alphabet based on size-extended benzo-homologated ts-bases can also form stable Watson-Crick base pairs with the natural complementary nucleobases.

1. Introduction

Emissive nucleobases have emerged as extremely useful biological tools for understanding at a molecular level the structure, activity, localization and interactions of nucleic acids.¹ Ideally, emissive (fluorescent) nucleobases are chemically modified nucleobases exhibiting improved fluorescence properties, while being isomorphic and isofunctional to their parent unmodified bases, i.e. retaining their chemical and biological functionalities, including base pairing and enzyme incorporation.²⁻⁵ As the natural nucleobases are inherently non-emissive, significant modifications are needed to provide them with emissive properties useful for a broad range of applications. Modifications that leave the Watson-Crick edge of the nucleobases, and relative H-bonding potential, unaffected are considered as “canonical”.¹ The first example of canonical fluorescent nucleobases dates back to 1969, when Stryer proposed the two emissive compounds formycin and 2-aminopurine (2AP).⁶ Since this seminal work, numerous fluorescent base analogues have been proposed and used in a variety of biotechnological and biomedical applications, as well as structural studies⁷⁻²³, including high-resolution imaging.²⁴

In this context, in 2015 Tor and coworkers synthesized a complete emissive RNA alphabet comprising isomorphic purine, ^{tz}A, ^{tz}G, and pyrimidine, ^{tz}U, ^{tz}C, analogues, all derived from isothiazolo[4,3-d]pyrimidine as a common heterocyclic core.²⁵ This was an evolution of their previously proposed alphabet derived from thieno[3,4-d]-pyrimidine.²¹ In both the alphabets the emissive nucleobases have their Watson-Crick edge untouched. However, in the latest proposed alphabet the purine analogues are more isomorphic and isofunctional to the natural bases, as they also preserve their N7 atom, whose basicity and coordinating ability is needed for biomolecular interactions.²⁵ Notably, the adenosine analogue ^{tz}A is indeed deaminated by adenosine deaminase as efficiently as the natural adenosine.²⁵

As this isothiazolo-based fluorescent alphabet seems to be extremely promising, we investigated it theoretically in a previous study, with a focus on the impact of the non-natural bases on the H-bonding properties and fluorescent properties.² Herein, we use it as a basis for the design of a novel alphabet, exhibiting enhanced fluorescence properties while preserving the isomorphism and possibly the isofunctionality of the original one. In this new alphabet, that we characterize here theoretically, we replaced the isothiazolo moiety with an isoselenazolo moiety, by substituting the sulfur atom at position 8 of the isothiazolo ring with a selenium (Se) atom.

The basis for such a substitution is in the fact that selenium is well known to feature anomalous diffraction in X-ray crystallography, thus helping solve the phase problem.^{26, 27} Furthermore, selenium is present in RNA natural posttranscriptional modifications²⁸ and has already been artificially introduced in nucleic acid bases, in place of an oxygen atom for solving the phase problem²⁹, but also for changing their pairing specificity^{30, 31}. Notably, these two effects have been combined in a recent study proposing a Se-based modification.³²

Figure 1. Structures of the natural (A/G/U/C), modeled fluorescent nucleobases (^{ts}A/ ^{ts}G/ ^{ts}C/ ^{ts}U) and the modeled size expanded non-natural nucleobases (^{xts}A/ ^{xts}G/ ^{xts}C/ ^{xts}U). The sugar-phosphate backbone of the natural and non-natural bases has been truncated at C1', notated as 'R' in the figure.

The nucleobases composing this new fluorescent alphabet are shown in Figure 1, we named them: ^{ts}A, ^{ts}G, ^{ts}C, and ^{ts}U (ts-bases). Herein, we also modeled a size-expanded RNA alphabet based on the ts-bases, where a benzene spacer is inserted between the isoselenazolo and the pyrimidine ring of the molecular scaffold. Nucleobases size-expansion by addition of an aromatic or heterocyclic ring is indeed a well-established strategy to increase their aromatic surface and, in turn, their emissive properties.^{11, 33} Very recently, a benzo-homologated (xtz-) isothiazole RNA alphabet has also been

proposed.³⁴ These bases are also shown in Figure 1, we named them: ^{xts}A, ^{xts}G, ^{xts}U, and ^{xts}C (xts-bases).

We performed electronic structure calculations to evaluate the impact of the modification on the structure and energy of the base pairs involving the Se-containing ts-modified nucleobases, and compared the results with those obtained for the S-containing tz-bases. In analogy to our previous studies,³⁵⁻⁴¹ we initially focused on the effect of the modification on the geometry and energetics of the H-bonded base pairs they participate in. First, we considered the classical *cis* Watson-Crick pairing (cWW), involving the unperturbed Watson-Crick edge, next, as the Hoogsteen edge in the two fluorescent purines also remains unmodified and available for H-bonding, we also investigated geometries involving the purines Hoogsteen edge in the same combinations we previously considered for the S-containing tz-bases². We characterized the structural and energetic impact of the modification on the Watson-Crick base pairing potential also for the size-expanded xts-bases.

In addition, we compared the aromatic character of the modeled ts-bases, as it can affect their stacking properties, and compared it with that of natural and tz-bases. Finally, we calculated the optical absorption properties of the modified nucleobases, compared them with those of natural and tz-bases, and rationalized them in terms of modification of the molecular orbitals. Our and other theoretical studies of fluorescent nucleobases have indeed shown that an accurate DFT protocol can fairly reproduce their experimental absorption and emission spectra.^{2, 18, 19} Overall, our calculations indicate that the sulfur-to-selenium substitution in the novel ts-bases causes enhanced absorption/emission properties, as compared to the tz-ones, while having only a minor effect on their H-bonding potential.

2. Models and Computational Details

2.1 Modeling the interaction systems. Initially, we focused on studying the impact on H-bonding of the modeled ^{ts}A, ^{ts}G, ^{ts}C, and ^{ts}U bases possessing the canonical Watson-Crick H-bonded base pairs in combination with the natural A, G, C, and U bases, by taking into account the ^{ts}A:U, A:^{ts}U, ^{ts}G:C, and G:^{ts}C base pairs. Then, we also studied the H-bonding interactions where both the bases were modified, that is the ^{ts}A:^{ts}U and ^{ts}G:^{ts}C base pairs. Similar to our previous study on tz-bases,² we further modeled four different H-bonding geometries, where modified purine bases, ^{ts}A and ^{ts}G, pairs utilizing their ‘Hoogsteen’ edge for H-bonding. The first three involve the participation of ^{ts}A,

utilizing its Hoogsteen edge to form H-bonding interaction with the Watson-Crick (W) edge of natural A/U/C bases, forming the ${}^{\text{ts}}\text{A}:\text{A}$, ${}^{\text{ts}}\text{A}:\text{U}$, ${}^{\text{ts}}\text{A}:\text{C}$ base pairs, all interacting with a *trans* orientation of the glycosidic bonds (tHW geometry). The fourth interaction we modeled is the ${}^{\text{ts}}\text{G}:\text{G}$ tHW base pair, where the modified ${}^{\text{ts}}\text{G}$ base uses its Hoogsteen edge to H-bond to the Watson-Crick edge of the natural G base. Next, we computationally modeled a size-expanded RNA alphabet based on the computationally modeled ts-bases. In particular, a benzene spacer was inserted between the isoselanzolo and the pyrimidine ring of the molecular scaffold, and the modified base analogs (xts-bases) were trivially named as ${}^{\text{xts}}\text{A}$, ${}^{\text{xts}}\text{G}$, ${}^{\text{xts}}\text{U}$, and ${}^{\text{xts}}\text{C}$, and their interaction with the natural A/U/G/C bases were studied, see Figure 4. For each of the modeled non-natural base pairs, we also modeled their corresponding unmodified/natural geometries. Thus, in total, we modeled twenty different combinations of the non-natural modified base pair systems along with their natural base pair counterparts. For all the model systems described above, the base pairs are truncated at the C1' atom of the ribose. This is the standard approach used in literature.^{36, 37, 40, 42-47}

2.2 Electronic structure calculations. The geometry optimizations of the H-bonded base pairs were performed utilizing a DFT approach, based on the hybrid B3LYP functional as implemented in the Gaussian09 package.⁴⁸⁻⁵⁰ The correlation-consistent polarized valence triple- ζ cc-pVTZ basis set⁵¹ was used for all the geometry optimizations in gas phase as well as in water, modeled with the C-PCM continuum solvation model.⁵² Since dispersion interactions might contribute differently to the stability of the base pairs under study, we explicitly added Grimme's D3 correction term to the electronic energy also geometry optimization.⁵³ Interaction energies were calculated at the second order Møller-Plesset (MP2)⁵⁴ level of theory, on the B3LYP-D3/cc-pVTZ optimized geometries, using the augmented aug-cc-pVTZ basis set. For these calculations we took advantage of the RIMP2⁵⁵ method as implemented in Turbomole 6.1 package, with water modeled with the continuum solvation model COSMO.⁵² All the interaction energies were corrected for the basis set super position error (BSSE),⁵⁶ using the standard counterpoise procedure.⁵⁶ Thus, the binding energy E_{Bind} is calculated as in Eq. 1:

$$E_{\text{Bind}} = E_{\text{Complex}} - (E_{\text{M1}} + E_{\text{M2}}), \quad (1)$$

where E_{Complex} is the electronic energy of the optimized M1:M2 base pair, E_{M1} and E_{M2} are the electronic energies of the optimized M1 and M2 bases. In addition, the basis set superposition error (BSSE) is removed from E_{Bind} using the counterpoise correction method proposed by Boys and Bernardi.⁵⁶

E_{Bind} of Eq. 1 can be rewritten as:

$$E_{\text{Bind}} = E_{\text{Int}} + E_{\text{Def}}, \quad (2)$$

Where E_{Def} is the deformation energy, corresponding to deform the isolated bases to the geometries they will assume in the complex, and E_{Int} is the interaction energy between these deformed geometries in the final complex.

This is a rather standard approach used in this kind of calculations.^{37, 40, 42-44, 57, 58} In the present study, we also derived the interaction energies in water, which were calculated using the same recipe as suggested by Sponer and coworkers.^{42, 59}

To have an immediate and intuitive understanding of the impact of a specific modification under study, we introduce the modification energy, E_{Mod} , defined as the energy difference between the binding energy of the modified and of the corresponding natural base pair (in this specific case A:U and G:C base pair),^{35, 36} as shown in Eq. 3.

$$E_{\text{Mod}} = E_{\text{Bind}}(\text{modified base pair}) - E_{\text{Bind}}(\text{natural base pair}). \quad (3)$$

Within this definition, positive and negative E_{Mod} values indicate modifications that decrease or increase the stability of a specific base pair, respectively, as E_{Bind} are, by definition, negative.

Absorption and emission spectra were calculated with time dependent density functional theory (TDDFT),⁶⁰⁻⁶² using the B3LYP functional as implemented in the Gaussian09 package. A similar theoretical approach has already been used on the non-natural emissive RNA alphabet derived from thieno[3,4-d]-pyrimidine and isothiazolo[4,3-d]pyrimidine core.^{2, 18, 19, 63} Geometry optimization of the singlet ground state (for absorption) and of the first singlet excited state (for emission) were performed without symmetry constraints. To calculate photophysical properties, we used the B3LYP functional, since a benchmark study we carried out to reproduce the experimental absorption and emission wavelengths of the tz-bases indicated it to perform better than other computational methods (see Table S1 of ESI).

Electron density analysis. Comparative analysis of the electron density of natural (A/U/G/C) and modified (^{ts}A, ^{ts}U, ^{ts}G, ^{ts}C) bases was performed as follows. First, we reoptimized the geometry of the ts-bases forcing C_s symmetry, with the symmetry plane coincident with the purine plane. Since all the nucleobases considered in this work are nearly planar, this constraint has no meaningful

impact on the structure/energy of the nucleobases, while facilitating the electron density difference analysis, which was performed in the symmetry plane.³⁸ Next we optimized the geometry of the natural bases by freezing the 6-membered rings to the positions they have in the corresponding ts -base. For instance, we replaced the C9 carbon of ts A with an N atom, plus we replaced the Se8 atom of ts A with a CH group to obtain the unmodified A base. The geometry of A was thus optimized by freezing the heavy atoms common to both the A and ts A skeleton (i.e., the N1, C2, N3, C4, C5, C6, N6 and N7 atoms). With this approach, the 6-atom rings and the N6 and N7 atoms of the ts A and the A bases can be perfectly superimposed, as shown in Figure S1, allowing safe comparison of the electron density differences for these two different edges between the modified and unmodified bases. Considering that Se is clearly larger than C, we did not fix the C1' atom and the heavy atoms in positions 8 and 9 of A. This approach results in a different geometries around positions 8 and 9, see Figure S1. A similar procedure was used for all the four ts -bases.

After optimization, we compared the RI-MP2/aug-cc-pVTZ electron densities of the modified ts A base, $\rho^{^{ts}A}$, and that of the unmodified A base with the geometry it has in the ts A base, $\rho^{A/^{ts}A}$. The identical geometries of the 6-membered rings avoid noise in the analysis of the electron density difference, $\rho^{^{ts}A-A/^{ts}A} = \rho^{^{ts}A} - \rho^{A/^{ts}A}$. A similar procedure was followed for the other bases. Similarly, to compare the electronic density differences between tz -base and ts -bases, we adopted the similar protocol.

Aromaticity analysis. Aromaticity of the nucleobases was quantified using the nucleus-independent chemical shifts (NICS) values proposed by Schleyer and co-workers,⁶⁴ computed using the gauge-including atomic orbital method⁶⁵ (GIAO) and including the solvent contribution by C-PCM. The NICS is defined as the negative of the absolute shielding computed at specific points around the molecular system. In particular, NICS(0) values are calculated at the ring centres, while NICS(d) values are calculated at a distance d (Å) above/below such ring centres.⁶⁶ In the current work, NICS(1) is defined as the average value of the shielding calculated 1.0 Å above and below the centre of mass of each ring in the bases. Thus, for purines, two NICS(1) values are reported, one corresponding to the 6-membered ring, one to the 5-membered ring.

3. Results and Discussion

In order to study the impact of the computationally designed fluorescent nucleobases, ts A, ts U, ts G and ts C (Figure 1), on the structure and stability of nucleic acids base pairs, 16 different base pair

models have been considered, half in the classical antiparallel Watson-Crick (cWW) geometry, characteristic of antiparallel duplex structures (see Figure 2), and the remaining ones in the *trans* Hoogsteen-Watson-Crick (tHW) geometry (see Figure 3), required for building several tertiary structural motifs.⁶⁷ Additional four base pairing geometries have been modeled for the size-expanded RNA alphabets (see Figures 1 and 4). The optimized geometry and the binding energy calculations were performed in the gas phase and in water for each of the modeled base pair. To have an immediate measure of the impact of a specific modification on the corresponding base pair, we used the modification energy, E_{Mod} , defined as the difference between the binding energy of the modified and of the corresponding unmodified base pair, for instance $^{\text{ts}}\text{G}:\text{C}$ cWW or $^{\text{ts}}\text{G}:^{\text{ts}}\text{C}$ cWW versus $\text{G}:\text{C}$ cWW (see Methods). By definition, negative and positive values of E_{Mod} indicate that the modified base pair is more or less stable than the corresponding unmodified base pair. Table 1 summarizes the calculated interaction energies and E_{Mod} values for the investigated base pairs. For comparison, the E_{Mod}^* have also been reported for the tz-base pairs, which have previously been calculated by us at the same level of theory.² Optimal geometries with the H-bonding distances in gas phase and in water are also reported for the base pairs forming ‘Watson-Crick’ and ‘Hoogsteen’ base pairs are shown in Figures 2 and 3, respectively. Similarly, the optimal gas phase and in water H-bonding distances of the ‘Watson-Crick’ geometries of the size-expanded xts-base pairs have been shown in Figure 4.

Table 1. Energy decomposition values, in kcal/mol. The cWW and tHW geometries for the non-natural ts-bases and their corresponding natural bases have been calculated. E_{Int} is the interaction energy of the two bases in the base pair without inclusion of deformation energy. E_{Def} is the deformation energy of the two bases calculated as the difference between the energy of the geometry they have in the base pair and the energy of the geometry they have when optimized alone. E_{Bind} is

the sum of the interaction and deformation energies, which is $E_{\text{Bind}} = E_{\text{Int}} + E_{\text{Def}}$. E_{Mod} is the difference between the binding energy of the modified base pair and of the reference unmodified pair. Negative and positive values of E_{Mod} indicate that the modified base pair is more stable or less stable than the corresponding unmodified base pair. $E^*\text{Mod}(\text{gas})$ and $E^*\text{Mod}(\text{W})$ corresponds to the modification energy of the tz-bases as calculated in reference.²

System	Geometry	E_{Int}	E_{Def}	$E_{\text{Bind}}(\text{gas})$	$E_{\text{Mod}}(\text{gas})$	$E^*_{\text{Mod}}(\text{gas})$	$E_{\text{Bind}}(\text{W})$	$E_{\text{mod}}(\text{W})$	$E^*_{\text{Mod}}(\text{W})$
A:U	cWW	-16.5	1.4	-15.0	-	-	-7.9	-	-
A: ^{ts} U	cWW	-16.3	1.4	-15.0	0.1	0.0	-7.9	0.0	0.0
^{ts} A:U	cWW	-16.7	1.5	-15.2	-0.2	-0.2	-8.0	-0.1	0.0
^{ts} A: ^{ts} U	cWW	-16.6	1.5	-15.1	0.0	0.0	-7.9	0.0	0.0
G:C	cWW	-30.7	2.8	-28.0	-	-	-12.5	-	-
G: ^{ts} C	cWW	-30.3	2.7	-27.6	0.4	0.4	-12.5	0.0	0.0
^{ts} G:C	cWW	-29.3	2.9	-26.4	0.6	0.4	-12.1	0.4	0.3
^{ts} G: ^{ts} C	cWW	-28.7	2.8	-26.0	2.0	0.9	-12.0	0.5	0.4
A:A	tHW	-13.8	1.3	-12.5	-	-	-7.1	-	-
^{ts} A:A	tHW	-12.5	1.6	-10.9	1.6	1.3	-6.5	0.7	0.7
A:U	tHW	-17.1	1.2	-15.9	-	-	-8.1	-	-
^{ts} A:U	tHW	-15.6	1.5	-14.1	1.8	2.7	-7.7	0.4	1.1
A:C	tHW	-17.1	1.6	-15.6	-	-	-7.3	-	-
^{ts} A:C	tHW	-16.0	1.9	-14.1	1.5	0.9	-6.6	0.7	0.7
G:G	tHW	-21.0	-0.1	-21.1	-	-	-9.0	-	-
^{ts} G:G	tHW	-20.9	0.5	-20.4	0.8	2.5	-9.1	-0.2	0.9

Figure 2. *Cis* Watson-Crick base pairs. Stick representation of unmodified and modified cWW A:U base pairs (left) and of unmodified and modified cWW G:C base pairs (right). Values in parentheses correspond to optimized H-bond distances in water and values out of parentheses correspond to optimized distances in the gas phase. All distances are in Å.

3.1 Structural and Energetic Impact of the modified nucleobases

Watson-Crick Base pairs. Eight base pairs featuring a cWW geometry have been considered, four of them corresponding to the modified/unmodified A:U pair and the other four to the modified/unmodified G:C pair (Figure 2). In two cases, ^{ts}A:^{ts}U and ^{ts}G:^{ts}C, the modifications are present on both the bases, while for the ^{ts}A:U/A:^{ts}U and ^{ts}G:C/G:^{ts}C pairs the modification is present on a single base. The optimized structures in gas phase and in water of the modified ^{ts}A:U, A:^{ts}U and ^{ts}A:^{ts}U cWW base pairs are extremely similar to that of the classical A:U cWW pair, with differences in H-bond lengths within 0.05 Å. The C1'-C1' distance, a parameter indicating the isostericity of H-bonded base pairs^{67, 68} is almost identical in the modified and unmodified A:U pairs (10.52 Å for A:U versus 10.48/10.56/10.51 Å for A:^{ts}U/^{ts}A:U/^{ts}A:^{ts}U). As for the base pairs, where the ^{ts}G and ^{ts}C modifications are included, a similar trend has been observed. The ^{ts}G:C, G:^{ts}C and ^{ts}G:^{ts}C base pair geometries are extremely similar to that of the natural G:C cWW base pair, with differences in H-

bond distances within 0.04 Å. Similarly, the C1'-C1' distances are also very similar (10.71 Å for G:C versus 10.82/10.74/10.82 Å for G:^{ts}C/^{ts}G:C/^{ts}G:^{ts}C). Moving on to the energetics, the binding energies of the modified pairs involving ^{ts}A and ^{ts}U bases are also extremely similar to that of the classical A:U cWW pair, with E_{Mod} within -0.2 kcal/mol both in the gas phase and in water. However, for the modified base pairs involving ^{ts}G and ^{ts}C, (see the ^{ts}G:^{ts}C pair), a destabilizing effect is found with E_{Mod} values within +2.0 and +0.4 kcal/mol in gas phase and in water, respectively. Overall, our geometric and energetic analysis indicates that the H-bonding properties on the 'Watson-Crick' edge are substantially conserved.

Non Watson-Crick Base pairs. Looking at the geometries of the ^{ts}A and ^{ts}G bases, it seems evident that their Hoogsteen edge remains unperturbed, allowing these non-natural purine bases to be engaged in potential non Watson-Crick H-bonding. Here, we have specifically studied the tHW ^{ts}A:A, ^{ts}A:U, ^{ts}A:C and ^{ts}G:G base pairs. In all these studied cases, the Hoogsteen edge of the modified ^{ts}A/^{ts}G bases interacts with the natural A/U/G/C bases utilizing its Watson-Crick edge, and in the *trans* orientation of glycosidic bonds. All the four systems were compared with the corresponding unmodified base pairs (see Figure 3).

The optimized geometries of all the modified tHW base pairs we analysed are quite similar to their corresponding unmodified base pairs. The H-bonding distance differences between the modified and unmodified geometries stay within 0.11 Å. We notice that the H-bond involving ^{ts}A/^{ts}G N7 is always elongated, while the other H-bond, involving the ^{ts}A/^{ts}G O6/N6 atoms is shortened. As for the binding energy, the modified pairs involving the ^{ts}A and ^{ts}G bases are relatively less stable than their corresponding unmodified pairs, see Table 1. The system that is most destabilized by the modification is the tHW ^{ts}A:U pair (E_{Mod} in gas phase of 1.8 kcal/mol), followed by the ^{ts}A:A pair (E_{Mod} in gas phase of 1.6 kcal/mol).

3.2 Structural and Energetic Comparison of ts-bases versus the tz-bases

The overall structures of the modeled ts-bases and the previously reported tz-bases seem to be isomorphic, with the only difference that the sulfur atom is replaced by a heavy selenium atom. The usage of a selenium atom in nucleic acid bases has already been reported.²⁹⁻³² The replacement of sulfur in tz-bases with selenium in ts-bases may possess two possible advantages, the former one being that the selenium in ts-bases could act as a scattering label for the phase determination in X-ray

crystallography, the latter one being the enhanced absorption and emission when ts-bases are used instead (discussed below, in Section 3.6).

Our previous computational investigation of tz-bases has clearly pointed out the minimal impact of these bases on the H-bonding and the energetics of the base pairs they are involved in.² Analogously, a very small impact on the H-bond distances of base pairs involving ts-bases has been observed here, both as compared to natural base pairs and to the tz-bases (H-bond distance deviations within 0.01 Å for the cWW base pairs and within 0.05 Å for tHW base pairs from the tz-bases). As for the energetics, the E_{Mod} of the ts-bases and the tz-bases are very similar for the cWW base pairs, apart from the $^{\text{ts}}\text{G}:^{\text{ts}}\text{C}$ base pair, where a destabilization of 1.1 kcal/mol has been calculated with respect to $^{\text{tz}}\text{G}:^{\text{tz}}\text{C}$ in the gas phase (see Table 1).² The low stability of this base pair may be correlated with the high deformation energy of 2.8 kcal/mol for $^{\text{ts}}\text{G}:^{\text{ts}}\text{C}$ versus 1.9 kcal/mol for the $^{\text{tz}}\text{G}:^{\text{tz}}\text{C}$ base pair. Next, for the tHW base pairs, a stabilization of 0.9 and 1.7 kcal/mol has been observed for the $^{\text{ts}}\text{A}:\text{U}$ and $^{\text{ts}}\text{G}:\text{G}$ pair, respectively, as compared to their corresponding tz-analogues.

Figure 3. Non-Watson Crick base pairs. Stick representation of optimized structures of the four tHW base pairs with a modified A/G using its Hoogsteen edge for H-bonding and corresponding unmodified pairs. Values in parentheses correspond to the optimized distances in water and values out of parentheses correspond to optimized distances in the gas phase. All distances are in Å.

Figure 4: Stick representation of the modified geometries of xts-bases paired with their natural complimentary bases. Values in parentheses correspond to optimized H-bond distances in water and values out of parentheses correspond to optimized distances in the gas phase. All distances are in Å.

3.3 Structure and Energetic Impact of Size-Expanded Modified base pairs involving xts-bases

The optimized geometries we obtained for the Watson-Crick (WC) base pairs formed between the xts-bases and their natural complementary bases are shown in Figure 4. The optimized base pairs exhibit planar geometries and possess similar H-bonding distances to the natural ones. In order to examine the influences introduced by the size-expansion on H-bonding ability, the interaction energies of these base pairs were also calculated. The calculated E_{Bind} are predicted to be -14.8/-15.4 and -27.9/-26.9 kcal/mol, respectively, for the A:^{xts}U/^{xts}A:U and the ^{xts}C:^{xts}G/^{xts}G:C base pairs. At the same level of theory, the calculated interaction energies for the corresponding natural and cWW base pairs were extremely similar, namely 15.0 for A:U and -28.0 kcal/mol for G:C. It means that the modified bases can form stable WC base pairs with natural complementary bases and the size-expansion would not influence greatly the pairing abilities of xts-bases with the natural bases, although the steric hindrance due to the additional benzene ring may impair the formation of a regular RNA duplex.

3.4 Electron density difference maps. The electron density difference maps of the ts-modified bases versus the unmodified/natural bases clearly show that changes in the electron density are negligible on the bases Watson-Crick edge, with minor electron density increase around the ^{ts}G O6, ^{ts}U O4 and ^{ts}C O2 and N4 atoms (see Figure 5A). This result is analogous to what we previously obtained for the tz-bases.² These observations are consistent with the comparable stability of the modified and unmodified base pairs, when involved in cWW geometries. As for the Hoogsteen edge, please notice that the comparison is limited to the two purine derivatives, as the additional ring of the modified pyrimidines makes comparison of the electron density on this edge meaningless. For the two fluorescent purine derivatives, a slight decrease in the electron density is observed on the N7 atom, as compared to the natural A/G bases, which reduces the basicity of N7 and may contribute to the decreased stability we observed for the base pairs where the Hoogsteen edge of a $^{ts}\text{G}/^{ts}\text{A}$ is involved in H-bonding (see section 3.1).

Finally, to get insight on how the electron density changes when sulfur of the tz-bases is replaced by selenium in the ts-bases, we also plotted the electron density difference between the ts- and tz-bases (see Figure 5B). A large increase in the electron density is obviously observed around the selenium atom. Atoms on the Watson-Crick edge are basically unaffected, while a slight increase in electronic density is observed around the N7 atom of the modified ts-bases, which suggests it to be a better H-bond acceptor in ts-bases than in the tz-bases. This may explain the shortening we observed in the H-bonds between N1(G) and N7(^{ts}G) in $^{ts}\text{G}:\text{G}$ tHW and between N3(U) and N7(^{ts}A) in $^{ts}\text{A}:\text{U}$ tHW, as compared to their tz-base pair counterparts (see section 3.1).²

Figure 5. Electron density maps. **A** Electron density differences, in the base plane, between modified and natural (unmodified) bases are shown on the structures of the four modified bases, $^{ts}\text{A/A}$, $^{ts}\text{G/G}$, $^{ts}\text{C/C}$ and $^{ts}\text{U/U}$. **B** Electron density differences, in the base plane, between ^{ts}A , ^{ts}G , ^{ts}C and ^{ts}U bases and ^{tz}A , ^{tz}G , ^{tz}C and ^{tz}U bases are shown on the structures of the four modified bases, $^{ts}\text{A}/^{tz}\text{A}$, $^{ts}\text{G}/^{tz}\text{G}$, $^{ts}\text{C}/^{tz}\text{C}$ and $^{ts}\text{U}/^{tz}\text{U}$. Density difference curves are plotted between -0.02 and 0.02 a.u., with a spacing of 0.001 a.u. Blue (red) lines refer to negative (positive) density difference curves, *i.e.*, to areas where the modified base presents reduced (increased) electron density, as compared to the corresponding natural base. Only atoms discussed in the text are labeled.

3.5 Aromaticity analysis In order to probe how the aromaticity of the 6- and 5-membered rings of the non-natural ts-bases is affected with respect to the tz-bases and their natural counterparts, the nucleus-independent chemical shift (NICS) descriptor was selected.⁶⁴ The ring is more aromatic in correspondence of larger values of the NICS, *i. e.* more negative values.^{69, 70} The NICS values of benzene are taken as reference. In addition, harmonic oscillator model of aromaticity (HOMA) values, defined as the normalized sums of squared deviations of bond lengths from the optimal values in benzene,⁷¹ were also calculated. All calculated NICS and HOMA values are reported in Table 2.

Comparison of the NICS(1) indices of natural and non-natural bases indicates that the aromaticity of the 6-membered ring of purine and pyrimidine bases is significantly reduced in the non-natural ts-bases. For instance, the NICS(1) value for the 6-membered ring of ^{ts}A is only -5.46 , versus -8.01 for A. A similar trend is observed for the other bases. In contrast, the aromaticity of the 5-membered ring of non-natural purines $^{ts}\text{A}/^{ts}\text{G}$, where the C8-H group in the natural base is replaced by a selenium atom in the non-natural ts-bases, is enhanced, although to a lower extent. For instance, the NICS(1) of the 5-membered ring of ^{ts}A is -10.80 , versus -9.71 for A. A similar trend is observed of

^{ts}G and G (see Table 2). This analysis clearly indicates that in the 6-membered ring of non-natural bases, electrons in the π -conjugated aromatic system are less delocalized relative to the natural nucleobase whereas the electron delocalization is more pronounced in the 5-membered ring of non-natural purines as compared to natural purines. The alternative HOMA analysis also supports the decrease of aromaticity of the 6-membered ring for the non-natural ^{ts}A, ^{ts}U, ^{ts}G, and ^{ts}C bases, when compared with the natural A, U, G and C bases. Finally, as for the comparison between the ts- and tz-bases, the NICS and HOMA indices are lower (in absolute value) for the modelled ts-bases, indicating a decreased aromaticity for these modified bases. This might have an influence on the stacking ability of the modified ts-nucleobases with the surrounding residues, to be further investigated.

Table 2. NICS(0), NICS(1) and HOMA aromaticity values for natural and modified ts-bases. For the sake of comparison, the previously reported values for tz-bases are also reported, in parenthesis.²

	6 membered ring			5 membered ring	
	NICS(0)	NICS(1)	HOMA	NICS(0)	NICS(1)
C ₆ H ₆	-8.227	-10.370	0.995	-	-
A	-6.032	-8.006	0.978	-11.201	-9.709
^{ts} A	-1.617(-2.538)	-5.461(-6.063)	0.664(0.757)	-12.496(-13.465)	-10.796(-11.630)
C	-1.347	-3.173	0.755	-	-
^{ts} C	+1.594(-0.800)	-1.655(-2.159)	0.467(0.625)	-11.911(-12.915)	-4.117(-10.470)
G	-2.909	-3.426	0.803	-10.611	-8.821
^{ts} G	+1.004(-0.081)	-1.669(-2.244)	0.326(0.549)	-11.416(-12.477)	-9.525(-10.474)
U	-1.089	-1.786	0.640	-	-
^{ts} U	+1.363(0.772)	-0.675(-0.945)	0.289(0.489)	-11.916(-12.881)	-9.612(-10.460)

3.6 Photophysical Properties of isoselenazolo[4,3-*d*]pyrimidine Nucleobase Analogues. The predicated photophysical properties of natural and modelled ts-bases nucleobases calculated in water using the TDDFT/B3LYP method are reported in Table 3. We find the lowest energy absorption peaks from the ground electronic state to the first excited singlet state at 351.2, 362.5, 350.6 and 329.6 nm for ^{ts}A, ^{ts}G, ^{ts}C and ^{ts}U, respectively. The highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of the studied ts-bases have the typical π molecular orbital characteristics, and we infer that the absorption and emission correspond to the π to π^* HOMO to LUMO transition only, from the ground state (S_0) to the first excited state (S_1). The

shapes of molecular orbitals for ts-bases and natural bases are given in Figure S2. For comparison, we also calculated the same absorption peaks for the natural nucleobases, 250.8, 248.4, 255.7 and 246.4 nm for A, G, C and U bases, versus experimental values of 259, 252, 271, and 262 nm, respectively.⁷² The comparison of absorption peaks for natural and ts-bases clearly indicates that modification in the ts-bases results in a bathochromic shift of roughly 100 nm and 114 nm for ^{ts}A and ^{ts}G, and of roughly 95 nm and 83 nm for ^{ts}C and ^{ts}U bases. Slightly larger bathochromic shifts were calculated in the emission spectra,⁷³ with bathochromic displacements of 130.6, 164.3, 133.1 and 119.6 nm for ^{ts}A, ^{ts}G, ^{ts}C and ^{ts}U, with respect to natural bases. In general, non-natural pyrimidine nucleobases show lower bathochromic shifts relative to the natural bases.

As explained in our earlier contribution², the origin of the bathochromic shift can be of course directly related to a reduced energy gap between the highest occupied and the lowest unoccupied molecular orbitals, the HOMO-LUMO gap. Focusing on energy, the data reported in Table 3 indicate that the energy of the HOMO is substantially unaltered by the modification, with differences between a non-natural base and its natural analogue within 0.2 eV. In contrast, the modification has a strong impact on the energy of the LUMO, which is roughly 1.0-1.6 eV lower in energy in the modified bases relative to the natural analogues. This reduces the HOMO-LUMO gap from roughly 5.3-5.5 eV in the natural bases to roughly 3.9-4.3 eV in the modified ts-bases, with the consequent bathochromic shift in the absorption and emission spectra. Remarkably, the respective bathochromic displacements of 18.9, 16.3, 17.8 and 17.0 nm for absorption and 6.8, 26.3, 23.4, 22.3 nm for emission were calculated for ^{ts}A, ^{ts}G, ^{ts}C and ^{ts}U compared to ^{tz}A, ^{tz}G, ^{tz}C and ^{tz}U bases, see Table 3. The comparative analysis of the HOMO and LUMO energy values for the natural, tz-bases and the ts-bases have been given in Figure 6. This analysis clearly indicates that newly modelled ts-bases are expected to exhibit better fluorescent properties when compared to previously reported tz-bases. Simulated absorption and emission spectra are reported in Figure S3.

Table 3. Predicted photophysical properties of ts-bases based on isoselenazolo[4,3-*d*]pyrimidine nucleobase analogues and HOMO (H), LUMO (L) and HOMO-LUMO gap energies (ΔE_{HL}). Values in parentheses correspond to the predicted absorption/emission (abs/em) values of the tz-bases;² os.

str. stands for the oscillator strength. The $\Delta(\text{abs})$, and $\Delta(\text{em})$ values correspond to the difference in absorption and emission between ts-bases and tz-bases.

Base	Absorption			Emission			H	L	ΔE_{HL}
	λ_{abs}	$\Delta(\text{abs})$	os. str.	λ_{em}	$\Delta(\text{em})$	os. str.			
^{ts} A	351.2(332.3)	18.9	0.165	429.4(422.6)	6.8	0.226	-6.25	-2.30	-3.95
^{ts} G	362.5(346.2)	16.3	0.127	455.9(429.6)	26.3	0.151	-6.01	-2.11	-3.90
^{ts} C	350.6(332.8)	17.8	0.157	441.1(417.7)	23.4	0.186	-6.34	-2.31	-4.03
^{ts} U	329.6(312.6)	17.0	0.149	402.5(380.2)	22.3	0.179	-6.62	-2.37	-4.25
A	250.8	-	0.249	298.78	-	0.449	-6.21	-0.85	-5.36
G	248.4	-	0.127	291.63	-	0.213	-5.84	-0.49	-5.35
C	255.7	-	0.119	307.97	-	0.082	-6.47	-1.00	-5.47
U	246.4	-	0.215	282.94	-	0.325	-6.83	-1.32	-5.51

Figure 6. Graphic representation of the $E(\text{eV})$ of the lowest unoccupied molecular orbitals, LUMOs, represented by L-nat for natural bases, L-tz for tz-bases and L-ts for ts-bases, and the highest occupied molecular orbitals, HOMOs, of the natural, tz-bases and ts-bases.

Conclusions

We have computationally characterized a new fluorescent RNA alphabet composed by four nucleobases, that we named ts-bases, derived by replacing sulfur with selenium in the previously

reported tz-bases.²⁵ From our analysis, the novel ^{ts}A, ^{ts}U, ^{ts}G, and ^{ts}C bases result to be highly isomorphic with the four natural nucleobases, although the additional isoselenazole ring on the pyrimidine derivatives might cause a steric hindrance in the context of RNA structures. We have investigated the H-bonding potential and the photophysical properties of these novel nucleobases. Our calculations indicate that these non-natural nucleobases have a minimal impact on the geometry and stability of classical Watson-Crick base pairs, as also confirmed by the insignificant difference in electron density between the Watson-Crick edges of non-natural and natural nucleobases. However, the non Watson-Crick base pairs involving the purines (^{ts}A, ^{ts}G) Hoogsteen edge get destabilized by up to 1.8 kcal/mol on modification as compared to their natural counterparts, suggesting that some RNA tertiary motifs could be slightly destabilized by these modified nucleobases. Analysis of bases aromaticity indicates reduced aromaticity for the 6-membered ring of non-natural purines and pyrimidines compared to the natural ones. In contrast, for the 5-membered ring of non-natural purines, ^{ts}A/^{ts}G, our analysis indicated an increased aromaticity relative to the natural bases, although at a lower extent. The impact of this modified aromaticity on the relative stacking interactions needs further investigation. Our calculations also show that the modification has a strong impact on the energy of the lowest unoccupied molecular orbitals (LUMO), which is roughly 1.0-1.6 eV lower in energy in the modified bases relative to the natural analogues, without affecting significantly the highest occupied molecular orbitals (HOMO). This results in a substantial reduction of the HOMO-LUMO gap in the modified bases, with the consequent bathochromic shift in the absorption and emission spectra. Remarkably, increased bathochromic displacement of 18.9, 16.3, 17.8 and 17.0 nm for absorption and 6.8, 26.3, 23.4, 22.3 nm for emission were calculated for ^{ts}A, ^{ts}G, ^{ts}C and ^{ts}U bases as compared to ^{tz}A, ^{tz}G, ^{tz}C and ^{tz}U bases, respectively. Overall, our analysis clearly indicates the newly modelled ts-bases as extremely promising fluorescent base analogues, in view of their expected better fluorescent properties as compared to the previously reported tz-bases, and proposes them for experimental investigation. Finally, we also show here that a size-expanded fluorescent RNA alphabet, where a benzene ring is inserted between the 5- and the 6-membered ring of the ts-bases, can form stable Watson-Crick base pairs with the natural complementary nucleobases.

ACKNOWLEDGEMENTS:

The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST). For computer time, this research used the resources of the

Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia. A.P. thanks the Spanish MINECO for a project CTQ2014-59832-JIN. We thank Professor Yitzhak Tor for helpful comments.

Supporting Information.

The SI consists of molecular orbitals calculated in the present work. Cartesian coordinates of all the structures and absorption and emission spectra, along with the photophysical properties calculated at different level of theory discussed in this work are included. This material is available free of charge via the Internet at <http://pubs.rsc.org>.

References:

- W. Xu, K. M. Chan and E. T. Kool, *Nat Chem*, 2017, **9**, 1043-1055.
- M. Chawla, A. Poater, R. Oliva and L. Cavallo, *Phys Chem Chem Phys*, 2016, **18**, 18045-18053.
- S. G. Srivatsan, N. J. Greco and Y. Tor, *Angew. Chem. Int. Ed.*, 2008, **47**, 6661-6665.
- S. G. Srivatsan and Y. Tor, *Tetrahedron*, 2007, **63**, 3601-3607.
- S. G. Srivatsan and Y. Tor, *Nat. Protoc.*, 2007, **2**, 1547-1555.
- D. C. Ward, E. Reich and L. Stryer, *J. Biol. Chem.*, 1969, **244**, 1228-1237.
- M. E. Hawkins, *Cell Biochem. Biophys.*, 2001, **34**, 257-281.
- M. E. Hawkins, W. Pfeleiderer, F. M. Balis, D. Porter and J. R. Knutson, *Anal. Biochem.*, 1997, **244**, 86-95.
- A. R. Hernandez and E. T. Kool, *Org. Lett.*, 2011, **13**, 676-679.
- A. R. Hernandez, L. W. Peterson and E. T. Kool, *ACS Chem. Biol.*, 2012, **7**, 1454-1461.
- A. T. Krueger and E. T. Kool, *J. Am. Chem. Soc.*, 2008, **130**, 3989-3999.
- A. H. Lee and E. T. Kool, *J. Am. Chem. Soc.*, 2006, **128**, 9219-9230.
- C. Liu and C. T. Martin, *J. Mol. Biol.*, 2001, **308**, 465-475.
- C. Liu and C. T. Martin, *J. Biol. Chem.*, 2002, **277**, 2725-2731.
- H. Liu, J. Gao, S. R. Lynch, Y. D. Saito, L. Maynard and E. T. Kool, *Science*, 2003, **302**, 868-871.
- H. Lu, K. He and E. T. Kool, *Angew. Chem. Int. Ed.*, 2004, **43**, 5834-5836.
- E. L. Rachofsky, E. Seibert, J. T. Stivers, R. Osman and J. B. Ross, *Biochemistry*, 2001, **40**, 957-967.
- P. K. Samanta, A. K. Manna and S. K. Pati, *Chem. Asian J.*, 2012, **7**, 2718-2728.
- P. K. Samanta, A. K. Manna and S. K. Pati, *J. Phys. Chem. B*, 2012, **116**, 7618-7626.
- D. I. Scopes, J. R. Barrio and N. J. Leonard, *Science*, 1977, **195**, 296-298.
- D. Shin, R. W. Sinkeldam and Y. Tor, *J. Am. Chem. Soc.*, 2011, **133**, 14912-14915.
- J. T. Stivers, *Nucleic Acids Res.*, 1998, **26**, 3837-3844.
- M. Chawla, I. Autiero, R. Oliva and L. Cavallo, *Phys Chem Chem Phys*, 2018, **20**, 3699-3709.
- G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates and X. W. Zhuang, *Nat. Methods*, 2011, **8**, 1027-.
- A. R. Rovira, A. Fin and Y. Tor, *J. Am. Chem. Soc.*, 2015, **137**, 14602-14605.
- M. S. Hunter, C. H. Yoon, H. DeMirici, R. G. Sierra, E. H. Dao, R. Ahmadi, F. Aksit, A. L. Aquila, H. Ciftci, S. Guillet, M. J. Hayes, T. J. Lane, M. Liang, U. Lundstrom, J. E. Koglin, P. Mgbam, Y. Rao, L. Zhang, S. Wakatsuki, J. M. Holton and S. Boutet, *Nat. Commun.*, 2016, **7**, 13388.
- P. S. Pallan and M. Egli, *Nat. Protoc.*, 2007, **2**, 647-651.
- P. Boccaletto, M. A. Machnicka, E. Purta, P. Piatkowski, B. Baginski, T. K. Wirecki, V. de Crecy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm and J. M. Bujnicki, *Nucleic Acids Res*, 2018, **46**, D303-D307.

29. L. N. Lin, J. Sheng and Z. Huang, *Chem. Soc. Rev.*, 2011, **40**, 4591-4602.
30. H. Sun, S. Jiang and Z. Huang, *Methods Mol. Biol.*, 2016, **1320**, 193-204.
31. H. Sun, J. Sheng, A. E. Hassan, S. Jiang, J. Gan and Z. Huang, *Nucleic Acids Res*, 2012, **40**, 5171-5179.
32. A. Nuthanakanti, M. A. Boerneke, T. Hermann and S. G. Srivatsan, *Angew. Chem. Int. Ed.*, 2017, **56**, 2640-2644.
33. A. T. Krueger, H. Lu, A. H. Lee and E. T. Kool, *Acc. Chem. Res.*, 2007, **40**, 141-150.
34. L. Zhang, X. Kong, M. Zheng and M. Wang, *J. Mol. Graph Model.*, 2017, **77**, 339-349.
35. M. Chawla, R. Credendino, R. Oliva and L. Cavallo, *J. Phys. Chem. B*, 2015, **119**, 12982-12989.
36. M. Chawla, R. Oliva, J. M. Bujnicki and L. Cavallo, *Nucleic Acids Res*, 2015, **43**, 9573.
37. R. Oliva, L. Cavallo and A. Tramontano, *Nucleic Acids Res*, 2006, **34**, 865-879.
38. R. Oliva, A. Tramontano and L. Cavallo, *RNA*, 2007, **13**, 1427-1436.
39. M. Chawla, R. Credendino, E. Chermak, R. Oliva and L. Cavallo, *J. Phys. Chem. B*, 2016, **120**, 2216-2224.
40. M. Chawla, S. Abdel-Azeim, R. Oliva and L. Cavallo, *Nucleic Acids Res.*, 2014, **42**, 714-726.
41. M. Chawla, R. Credendino, A. Poater, R. Oliva and L. Cavallo, *J. Am. Chem. Soc.*, 2015, **137**, 299-306.
42. J. E. Sponer, K. Reblova, A. Mokdad, V. Sychrovsky, J. Leszczynski and J. Sponer, *J. Phys. Chem. B*, 2007, **111**, 9153-9164.
43. J. E. Sponer, N. Spackova, P. Kulhanek, J. Leszczynski and J. Sponer, *J. Phys. Chem. A*, 2005, **109**, 2292-2301.
44. J. E. Sponer, N. Spackova, J. Leszczynski and J. Sponer, *J. Phys. Chem. B*, 2005, **109**, 11399-11410.
45. P. Sharma, L. A. Lait and S. D. Wetmore, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2435-2448.
46. M. Chawla, E. Chermak, Q. Zhang, J. M. Bujnicki, R. Oliva and L. Cavallo, *Nucleic Acids Res.*, 2017, **45**, 11019-11032.
47. M. Chawla, R. Oliva, J. M. Bujnicki and L. Cavallo, *Nucleic Acids Res.*, 2015, **43**, 6714-6729.
48. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652.
49. A. D. Becke, *Abstracts of Papers of the American Chemical Society*, 1996, 212, 112-COMP.
50. C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.
51. T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007-1023.
52. A. Klamt and G. Schuurmann, *Journal of the Chemical Society-Perkin Transactions 2*, 1993, DOI: Doi 10.1039/P29930000799, 799-805.
53. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, 132.
54. C. Moller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 0618-0622.
55. F. Weigend and M. Haser, *Theo. Chem. Acc.*, 1997, **97**, 331-340.
56. S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553-.
57. P. Sharma, S. Sharma, M. Chawla and A. Mitra, *J Mol Model*, 2009, **15**, 633-649.
58. M. Chawla, P. Sharma, S. Halder, D. Bhattacharyya and A. Mitra, *J Phys Chem B*, 2011, **115**, 1469-1484.
59. C. L. Zirbel, J. E. Sponer, J. Sponer, J. Stombaugh and N. B. Leontis, *Nucleic Acids Res*, 2009, **37**, 4898-4918.
60. R. Bauernschmitt and R. Ahlrichs, *Chemical Physics Letters*, 1996, **256**, 454-464.
61. M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, *J. Chem. Phys.*, 1998, **108**, 4439-4449.
62. R. E. Stratmann, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 1998, **109**, 8218-8224.
63. M. Gedik and A. Brown, *Journal of Photochemistry and Photobiology a-Chemistry*, 2013, **259**, 25-32.
64. P. V. Schleyer, C. Maerker, A. Dransfeld, H. J. Jiao and N. J. R. V. Hommes, *J. Am. Chem. Soc.*, 1996, **118**, 6317-6318.
65. K. Wolinski, J. F. Hinton and P. Pulay, *J. Am. Chem. Soc.*, 1990, **112**, 8251-8260.
66. J. Poater, M. Duran, M. Sola and B. Silvi, *Chem. Rev.*, 2005, **105**, 3911-3947.
67. N. B. Leontis, J. Stombaugh and E. Westhof, *Nucleic Acids Res.*, 2002, **30**, 3497-3531.
68. N. B. Leontis and E. Westhof, *RNA*, 2001, **7**, 499-512.

69. C. Corminboeuf, T. Heine, G. Seifert, P. V. Schleyer and J. Weber, *Phys. Chem. Chem. Phys.*, 2004, **6**, 273-276.
70. P. V. Schleyer, M. Manoharan, H. J. Jiao and F. Stahl, *Org. Lett.*, 2001, **3**, 3643-3646.
71. T. M. Krygowski and M. K. Cyranski, *Chem. Rev.*, 2001, **101**, 1385-1419.
72. M. J. Cavalluzzi and P. N. Borer, *Nucleic Acids Res.*, 2004, **32**, e13.
73. D. Onidas, D. Markovitsi, S. Marguet, A. Sharonov and T. Gustavsson, *J. Phys. Chem. B*, 2002, **106**, 11367-11374.

Material for the Table of Contents