



Cite this: DOI: 10.1039/c8ob01669k

Received 13th July 2018,
 Accepted 22nd August 2018
 DOI: 10.1039/c8ob01669k
 rsc.li/obc

Why do A-T and G-C self-sort? Hückel aromaticity as a driving force for electronic complementarity in base pairing†

Yu Zhang, Chia-Hua Wu and Judy I-Chia Wu *

Density functional theory computations and block-localized wavefunction analyses for 57 hydrogen-bonded base pairs document excellent linear correlation between the gas-phase association energies and the degree of aromaticity gain of paired bases ($r = 0.949$), challenging prevailing views of factors that underlie the proposed electronic complementarity of A-T(U) and G-C base pairs. Base pairing interactions can polarize the π -electrons of interacting bases to increase (or decrease) cyclic $4n + 2\pi$ electron delocalization, resulting in aromaticity gain (or loss) in the paired bases, and become strengthened (or weakened). The potential implications of this reciprocal relationship for improving nucleic acid force-fields and for designing robust unnatural base pairs are discussed.

Introduction

More than sixty years have passed since the proposal of the double helix structure of DNA,¹ yet fundamental aspects of the recognition properties of nucleobase pairs remain puzzling. How does Nature choose the optimal hydrogen bonding complement for a specific nucleobase (and can we mimic this selectivity)? Given a mixture of adenine (A), thymine (T)/uracil (U), guanine (G), and cytosine (C) in the primordial soup, why does A pair with T (or U) and G with C instead of to themselves? In this work, we report computational evidence suggesting that aromaticity gain (or loss) in paired bases can strengthen (or weaken) base pairing interactions, having direct relevance for rationalizing the electronic complementarity of A-T(U) and G-C pairs in DNA and RNA and for designing unnatural hydrogen-bonded base pairs.

In their seminal work, Kyogoku, Lord, and Rich first evoked the attractive idea that the A-T(U) and G-C pairs might exhibit special electronic features, *i.e.*, “electronic complementarity”, favoring their specific associations.^{2,3} Measurements of the association constants (K_{assoc}) of these nucleobases and their derivatives in chloroform revealed noticeably higher K_{assoc} values for the A-U (100 M^{-1}) pair, compared to A-A ($\sim 3 \text{ M}^{-1}$) and U-U ($\sim 6 \text{ M}^{-1}$), and the G-C ($10^4\text{--}10^5 \text{ M}^{-1}$) pair, compared to G-G ($10^3\text{--}10^4 \text{ M}^{-1}$) and C-C ($\sim 28 \text{ M}^{-1}$).^{2,3} The recognition of A-U caught special attention since the self-associated A-A and

U-U also formed two hydrogen bonds. It was proposed that the A-U pair might exhibit additional attractive C-H…O interactions between the H8 of A and the O2 of U (Fig. 1a).^{4,5} Others pointed out, however, that in both the Watson-Crick and Hoogsteen configurations of A-U, the C-H…O interactions were distal, nonlinear, and thus at most weak interactions.⁶⁻⁹

Here, we show that the aromatic characters of nucleobases (*i.e.*, their “ π -conjugation patterns”) influence their association strengths to complementary bases through a reciprocal aromaticity-modulated hydrogen bonding (AMHB) relationship.^{10,11} Base pairing interactions that increase aromaticity (*i.e.*, enhance cyclic $4n + 2\pi$ -electron delocalizations) of the interacting bases exhibit stronger than expected hydrogen bonds, while those that decrease aromaticity (*i.e.*, disrupt cyclic $4n + 2\pi$ -electron delocalizations) of the interacting bases display weaker associations. In a related work, Cyrański *et al.* showed indeed that hydrogen bonding at the C=O positions of T, G, and C base pairs increased the aromatic characters of the respective rings.¹² Fliegl *et al.* reported that the interaction strengths of several hydrogen-bonded dimers, including the Watson-Crick, A-T and G-C pairs, correlated to their computed diamagnetic susceptibilities.¹⁴ Energy decomposition analyses for A-T and G-C quantified the effects of resonance-assistance.^{7,13} Demonstrative examples of AMHB, in squaramide complexes^{15,16} and polymers,¹⁷ in dimers of five and six membered arrays,^{10,11} and in multipoint hydrogen bonded arrays¹⁸ also have been reported.

Schematic illustrations of aromaticity-modulated hydrogen bonding in the A-T and G-C base pairs are shown Fig. 1. In both the Watson-Crick and natural Hoogsteen configurations of A-T (Fig. 1a and b), hydrogen bonding interactions polarize

Department of Chemistry, University of Houston, Houston, TX, 77204, USA.

E-mail: jiwu@central.uh.edu

† Electronic supplementary information (ESI) available. See DOI: [10.1039/c8ob01669k](https://doi.org/10.1039/c8ob01669k)

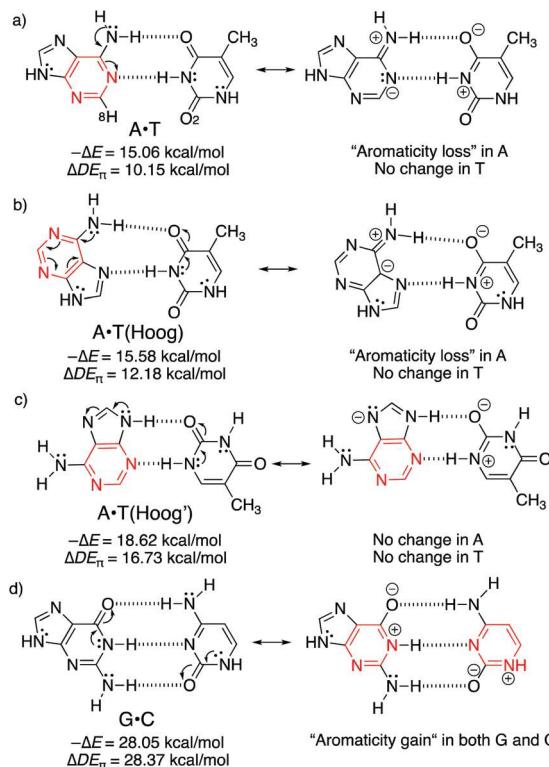


Fig. 1 Aromaticity-modulated hydrogen bonding (AMHB) in the (a) Watson–Crick A·T, (b) natural Hoogsteen A·T, (c) most stable Hoogsteen A·T, and (d) Watson–Crick G·C pairs. Resonance structures with formal cyclic $4n + 2\pi$ electron delocalizations are in red. Computed interaction energies ($-\Delta E$) and the estimated π -conjugation gain ($-\Delta E_{\pi}$) effects also are shown.

the ring π -electrons of the bases modestly, leading to decreased aromatic character in A, while T remains non-aromatic. In the most stable A·T configuration, A·T(Hoog')

(Fig. 1c), hydrogen bonding interactions polarize the ring π -electrons, but result in no gain or loss of aromatic character in either base. In the Watson–Crick G·C pair (Fig. 1d), hydrogen bonding interactions polarize the ring π -electrons of both G and C, leading to increased aromatic character in both bases (note resonance form in red), and the resulting "aromaticity gain" stabilizes the G·C complex in addition to the three hydrogen bonds present. We show that in this way, base pairs with the same numbers and types of hydrogen bonds can exhibit notably different pairing strengths depending on the π -conjugation pattern of the base.

Results and discussion

Based on a survey of 57 natural and unnatural base pairs, excellent linear correlation ($r = 0.949$, Fig. 2) was found between the gas-phase association energies of each base pair (a·b) ($\Delta E = E_{a·b} - E_a - E_b$) and the propensity of the interacting bases to gain or lose aromatic character ($-\Delta E_{\pi}$, see below). Geometries for all structures were optimized with a constrained C_s symmetry at ω B97X-D/6-311+G(d,p) employing Gaussian09¹⁹ (see details in the ESI†). Base pairs subject to obvious steric effects were excluded from the study.

Since aromaticity is related to the degree of π -electron delocalization in molecules, the effects of aromaticity gain or loss can be quantified by the amount of increase in π -electron delocalization upon base pairing, and is evaluated here by the block-localized wavefunction (BLW) analysis.^{20–22} BLW quantified the π -electron delocalization energy (DE_{π}) of the base pairs and bases by comparing the fully delocalized wavefunction (ψ_{deloc}) of the system considered to that of a hypothetical localized wavefunction (ψ_{loc}), in which all π -electrons were

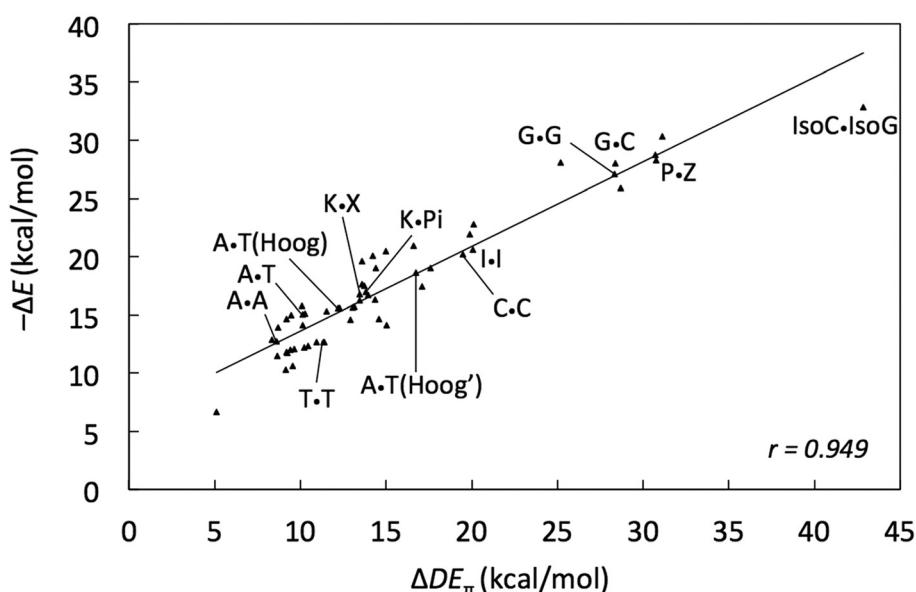


Fig. 2 Plot of base pairing interaction energy ($-\Delta E$, in kcal mol^{-1}) vs. π -conjugation gain (ΔE_{π}) in the gas-phase for all 57 base pairs. Plot of $-\Delta E$ vs. ΔE_{π} for selected base pairs in chloroform is provided in Fig. S8 of the ESI.†

mathematically constrained to resemble a strict π -electron-localized Lewis structure; $DE_{\pi} = \psi_{\text{loc}} - \psi_{\text{deloc}}$. The increase in π -electron delocalization energy (ΔDE_{π}) (as a result of base pairing) is evaluated by the computed DE_{π} value for the base pair considered (a-b) minus that of the interacting bases (a and b); $\Delta DE_{\pi} = DE_{\text{a-b}} - (DE_{\text{a}} + DE_{\text{b}})$ (see details in the ESI[†]). All BLW computations were performed at B3LYP/6-31G(d) employing the GAMESS-2013-R1 program.²³

Following this procedure, the computed ΔDE_{π} values for all 57 base pairs were positive, indicating increased π -conjugation for all paired bases upon hydrogen bonding. The amount of π -conjugation gain differs depending on whether there is an increase or decrease in aromatic character in the paired bases. Higher ΔDE_{π} values indicate more aromaticity gain upon base pairing; lower ΔDE_{π} values indicate little to no aromaticity gain or aromaticity loss. For example, the computed ΔDE_{π} values for the Watson-Crick and natural Hoogsteen A-T pairs (10.2 and 12.2 kcal mol⁻¹, aromaticity loss in A, no change in T, Fig. 1a and b) are lower compared to that of the most stable A-T configuration, A-T(Hoog²), (16.7 kcal mol⁻¹, no change in aromaticity for A or T, Fig. 1c). The computed ΔDE_{π} for G-C (28.4 kcal mol⁻¹) is even higher since base pairing increases aromaticity in both G and C (Fig. 1a).

Accordingly, the computed electrostatic potential (ΔESP) difference maps for the Watson-Crick, A-T and G-C, pairs show stark differences, indicating very different polarizabilities for A, T, G, and C (Fig. 3). The ΔESP plots of A and T (upon pairing to form A-T) showed relatively little electron polarization in the ring moieties, while those of G and C (in G-C) showed notable polarization in the ring. Positive ΔESP values (blue) indicate a more repulsive surface, and negative ΔESP values (red) a more attractive surface upon base pairing. Each plot was generated by comparing the computed ESP values of the paired bases minus that of the isolated bases at a 0.001 a.u.

isosurface (generated by the Multiwfn program,^{24,25} see details in the ESI[†]). We note that previous benchmarking studies of the performance of various force-fields²⁶ against quantum mechanical methods documented better agreement for the computed interaction energies of base pairs such as A-T, A-A, and T-T (aromaticity loss or no change), relative to base pairs such as G-C and G-G (aromaticity gain). It is tempting to make the connection that such variations, *i.e.*, differences in the polarizability of nucleobases because of their π -conjugation patterns, may explain why fixed-charged approaches adopted by popular force-fields,^{27,28} might understabilize certain interactions but overstabilize others.

Considering the potential for aromaticity gain or loss in base pairs could help explain variations in their association strengths. For example, it has been suggested that, among the doubly hydrogen-bonded, self-associated, G-G, C-C, T-T, A-A pairs, G-G and C-C displayed especially high association strengths due to additional attractive secondary electrostatic interactions (SEI);²⁹ in G-G, between the amino groups on C2 and the carbonyl groups on C6, and in C-C, between the amino groups on C4 and the carbonyl groups on C2. In T-T, there are additional repulsive SEI's between the C2 and C4 C=O groups. These attractive interactions are absent in A-A. More recent studies suggested the important effects of steric repulsion on base pairing in G-G vs. C-C.³⁰ We show here that, in addition to the SEI and possible steric effects, the strong association of G-G (as well as its closely related inosine analog, I-I) may be attributed to prospects for significant aromaticity gain in the paired G (and I) bases; note the aza-2-pyridone moieties of G-G and I-I (Fig. 4). In C-C and T-T, base pairing has little to no effect on the aromatic character of either monomer. In A-A, base pairing reduces the aromatic character of the paired A units; note the 2-hydroxypyridine moiety of A-A (Fig. 4). Relevant resonance forms are shown in Fig. S2 of the ESI[†].

Direct comparisons of the computed $-\Delta E$ values for G-G, I-I, C-C, T-T, A-A, to those of their hydrogen-bonded acyclic dimer references (1-1, 2-2, 3-3, 4-4, 5-5) document the energetic effects of AMHB (Fig. 4). Notably, the computed $-\Delta E$ values for G-G (27.1 kcal mol⁻¹) and I-I (20.6 kcal mol⁻¹) are 6 to 8 kcal mol⁻¹ higher compared to those of their acyclic references, 1-1 (19.6 kcal mol⁻¹) and 2-2 (14.3 kcal mol⁻¹), which display the same primary and secondary electrostatic interactions but are preclude of aromaticity gain. In contrast, the computed $-\Delta E$ values for C-C (20.2 kcal mol⁻¹) and T-T (12.7 kcal mol⁻¹) closely follow those of their acyclic references, 3-3 (21.1 kcal mol⁻¹) and 4-4 (10.8 kcal mol⁻¹), suggesting that key factors relevant to the hydrogen bond strengths of C-C and T-T are adequately captured by their acyclic references. The computed $-\Delta E$ for A-A (12.8 kcal mol⁻¹) is modestly lower than 5-5 (14.4 kcal mol⁻¹), as expected by aromaticity loss of A upon base pairing.

Recognizing the effect of AMHB also has important implications for synthetic efforts in “expanding the genetic alphabet”. Several research groups have demonstrated elegant examples of artificial replication processes mimicking DNA, by

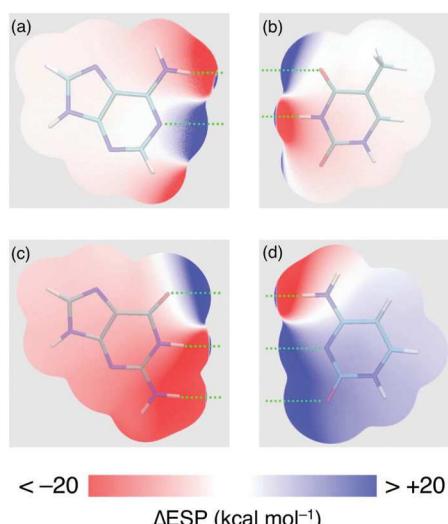


Fig. 3 Computed electrostatic potential difference maps, ΔESP , for (a) adenine, (b) thymine (c) guanine, and (d) cytosine, upon base pairing to A-T and G-C.

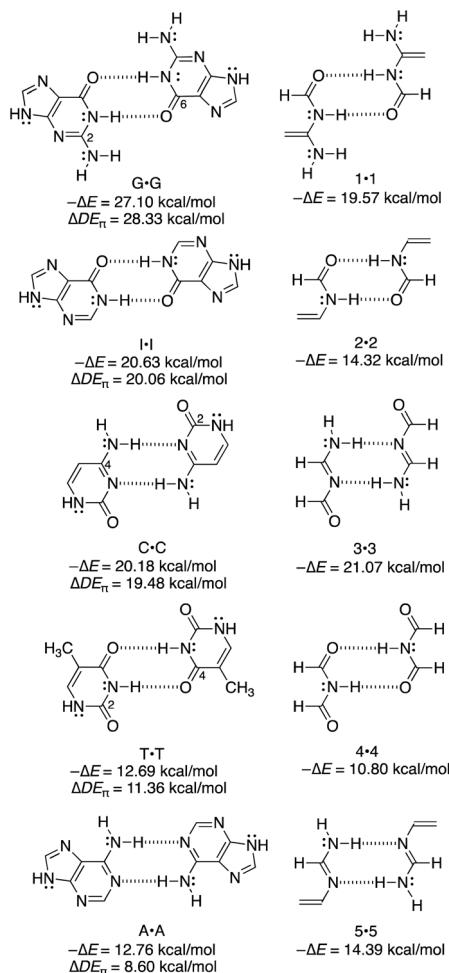


Fig. 4 Computed $-\Delta E$ and ΔDE_{π} values for the self-associated G·G, I·I, C·C, U·U, A·A pairs, and $-\Delta E$ values for their acyclic references, 1·1, 2·2, 3·3, 4·4, and 5·5. See also Fig. S2 in the ESI.†

using “unnatural” base pairs.^{31–34} Although the designs of unnatural base pairs have focused primarily on optimizing geometric complementarity (in which hydrogen bonds may or may not be present), the correlation shown in Fig. 2 suggests, that for hydrogen-bonded pairs, aromaticity gain (and loss) may serve as an effective strategy for modulating the robustness of unnatural base pairs, such as the isoC·isoG, P·Z, K·Pi, K·X pairs discussed below.

As shown in Fig. 5, the computed $-\Delta E$ values for both isoC·isoG (32.9 kcal mol^{−1}) and P·Z (28.3 kcal mol^{−1}) are 5 to 10 kcal mol^{−1} higher than their acyclic reference 3·1 (22.9 kcal mol^{−1}), due to increased aromaticity in the isoC, isoG, P, Z moieties upon base pairing. In sharp contrast, the computed $-\Delta E$ values for both K·Pi (17.0 kcal mol^{−1}) and K·X (16.8 kcal mol^{−1}) are close to that of their acyclic reference 6·4 (15.8 kcal mol^{−1}), indicating little non-additivity beyond the primary and secondary electrostatic effects present (base pairing decreases the aromatic character of K, and has little to no effect on the aromatic character of Pi and X). Relevant resonance forms are shown in Fig. S3 of the ESI.† A plot

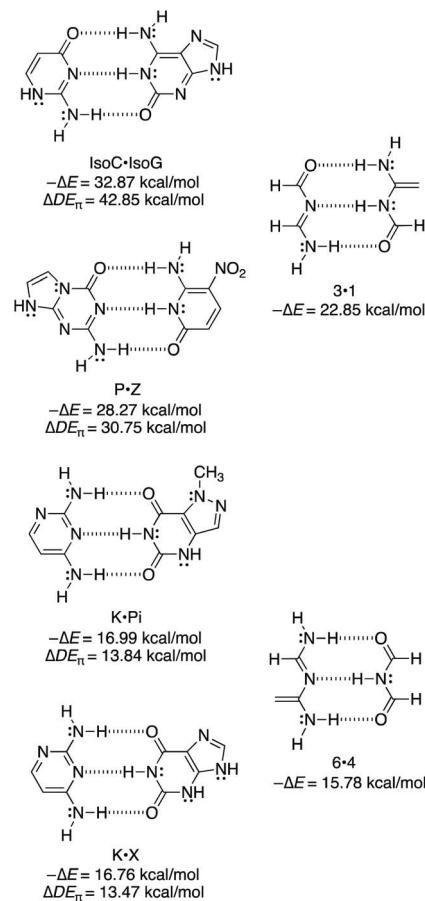


Fig. 5 Computed $-\Delta E$ and ΔDE_{π} values for isoC·isoG, P·Z, K·Pi, K·X, and $-\Delta E$ values of their acyclic references. See also Fig. S3 in the ESI.†

showing linear correlation, between $-\Delta E$ vs. ΔDE_{π} , for 1·1, 2·2, 3·3, 4·4, 5·5, 3·1, 6·4 is provided in Fig. S9 of the ESI.†

Overall, our findings suggest that while primary and secondary electrostatic interactions²⁹ have clear energetic consequences for base pairing (e.g., $-\Delta\Delta E = 8.8$ kcal mol^{−1} for 1·1 vs. 4·4, and 7.1 kcal mol^{−1} for 3·1 vs. 6·4), the effects of AMHB are comparable in magnitude (e.g., $-\Delta\Delta E = 7.5$ kcal mol^{−1} for 1·1 vs. G·G, and 10.0 kcal mol^{−1} for 3·1 vs. isoC·isoG), and therefore should be considered when evaluating base pairing strengths.

Conclusions

It is perhaps curious that adenine is the only fully “aromatic” nucleobase in the genetic code according to the Hückel $4n + 2\pi$ electron rule for aromaticity. None of the other bases in DNA or RNA, *i.e.*, thymine, uracil, cytosine, guanine, inosine, are $4n + 2\pi$ electron “aromatic”, despite having a closed-shell, cyclic, π -conjugated structure. What emerges from our finding is the suggested possibility that the π -conjugation patterns encoded to nucleobases have real chemical significance for modulating, understanding, and perhaps simulating base pairing interactions in DNA and RNA.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the National Science Foundation for grant support (CHE-1751370), as well as computational resources provided by the uHPC cluster, managed by the University of Houston and acquired through support from the NSF (MRI-1531814).

Notes and references

- 1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
- 2 Y. Kyogoku, R. C. Lord and A. Rich, *Biochim. Biophys. Acta*, 1969, **179**, 10–17.
- 3 Y. Kyogoku, R. C. Lord and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1967, **57**, 250–257.
- 4 K. S. Jeong, T. Tjivikua, A. Muehldorf, G. Deslongchamps, M. Famulok and J. Rebek Jr., *J. Am. Chem. Soc.*, 1991, **113**, 201–209.
- 5 G. A. Leonard, K. McAuley-Hecht, T. Brown and W. N. Hunter, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1995, **51**, 136–139.
- 6 J. R. Quinn, S. C. Zimmerman, J. E. Del Bene and I. Shavitt, *J. Am. Chem. Soc.*, 2007, **129**, 934–941.
- 7 C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders and E. J. Baerends, *Chem. – Eur. J.*, 1999, **5**, 3581–3594.
- 8 A. Asensio, N. Kobko and J. J. Dannenberg, *J. Phys. Chem. A*, 2003, **107**, 6441–6443.
- 9 P. Hobza, J. Sponer, E. Cubero, M. Orozco and F. J. Luque, *J. Phys. Chem. B*, 2000, **104**, 6286–6292.
- 10 J. I. Wu, J. E. Jackson and P. v. R. Schleyer, *J. Am. Chem. Soc.*, 2014, **136**, 13526–13529.
- 11 T. Kakeshpour, J. I. Wu and J. E. Jackson, *J. Am. Chem. Soc.*, 2016, **138**, 3427–3432.
- 12 M. K. Cyrański, M. Gilski, M. Jaskólski and T. M. Krygowski, *J. Org. Chem.*, 2003, **68**, 8607–8613.
- 13 J. F. Beck and Y. Mo, *J. Comput. Chem.*, 2006, **28**, 455–466.
- 14 H. Fliegl, O. Lehtonen, D. Sundholm and V. R. I. Kaila, *Phys. Chem. Chem. Phys.*, 2011, **13**, 434–437.
- 15 D. Quiñonero, A. Frontera, P. Ballester and P. M. Deyà, *Tetrahedron Lett.*, 2000, **41**, 2001–2005.
- 16 D. Quiñonero, R. Prohens, C. Garau, A. Frontera, P. Ballester, A. Costa and P. M. Deyà, *Chem. Phys. Lett.*, 2002, **351**, 115–120.
- 17 V. S. Talens, P. Englebienne, T. T. Trinh, W. E. M. Noteborn, I. K. Voets and R. E. Kieltyka, *Angew. Chem.*, 2015, **127**, 10648–10652.
- 18 C.-H. Wu, Y. Zhang, K. v. Rickley and J. I. Wu, *Chem. Commun.*, 2018, **54**, 3512–3515.
- 19 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. A. Montgomery Jr., 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 and D. J. Fox, *Gaussian 09, revision D.01*, Gaussian, Inc., Wallingford, CT, 2013.
- 20 Y. Mo, J. Gao and S. D. Peyerimhoff, *J. Chem. Phys.*, 2000, **112**, 5530–5538.
- 21 Y. Mo, L. Song and Y. Lin, *J. Phys. Chem. A*, 2007, **111**, 8291–9301.
- 22 Y. Mo, in *The Chemical Bond: Fundamental Aspects of Chemical Bonding*, ed. G. Frenking and S. Shaik, Wiley, Weinheim, Germany, 2014, p. 199.
- 23 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347–1363.
- 24 T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580–592.
- 25 T. Lu and F. Chen, *J. Mol. Graphics Modell.*, 2012, **38**, 314–323.
- 26 P. Hobza, M. Kabelac, J. Sponer, P. Mejzlik and J. Vondrasek, *J. Comput. Chem.*, 1996, **18**, 1136–1150.
- 27 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- 28 D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
- 29 W. L. Jorgensen and J. Pranata, *J. Am. Chem. Soc.*, 1990, **112**, 2008–2010.
- 30 S. C. C. van der Lubbe and C. Fonseca Guerra, *Chem. – Eur. J.*, 2017, **23**, 10249–10253.
- 31 E. T. Kool, *Acc. Chem. Res.*, 2002, **35**, 936–943.
- 32 J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, *Nature*, 1990, **343**, 33–37.
- 33 M. Ishikawa, I. Hirao and S. Yokoyama, *Tetrahedron Lett.*, 2000, **41**, 3931–3934.
- 34 T. Ohtsuki, M. Kimoto, M. Ishikawa, T. Mitsui, I. Hirao and S. Yokoyama, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4922–4925.